

Type	L #	Hits	Search Text	DBs	Time Stamp	Comments	Error Definition	Errors
1 BRS	L28	122	staphylococcus adj aureus	USPAT; US-PGPUB; EPO; JPO; DERWENT	2004/01/16 16:29			0
2 BRS	L29	0	s. adj aureus	USPAT; US-PGPUB; EPO; JPO; DERWENT	2004/01/16 16:29			0
3 BRS	L30	2	dnai adj polypeptide	USPAT; US-PGPUB; EPO; JPO; DERWENT	2004/01/16 16:30			0
4 BRS	L31	24	dnai	USPAT; US-PGPUB; EPO; JPO; DERWENT	2004/01/16 16:30			0
5 BRS	L33	0	28 same (30 or 31)	USPAT; US-PGPUB; EPO; JPO; DERWENT	2004/01/16 16:33			0
6 BRS	L34	2	31 same binding	USPAT; US-PGPUB; EPO; JPO; DERWENT	2004/01/16 16:33			0
7 BRS	L35	1	inhibitor same 34	USPAT; US-PGPUB; EPO; JPO; DERWENT	2004/01/16 16:34			0
8 BRS	L36	46988	bacteri\$2 same infection	USPAT; US-PGPUB; EPO; JPO; DERWENT	2004/01/16 16:35			0
9 BRS	L37	2	36 same 30	USPAT; US-PGPUB; EPO; JPO; DERWENT	2004/01/16 16:38			0
10 BRS	L38	1	31 same inhibitor	USPAT; US-PGPUB; EPO; JPO; DERWENT	2004/01/16 16:39			0

THIS PAGE BLANK (USPTO)

	Type	L #	Hits	Search Text	DBs	Time Stamp	Comments	Error Definition	Errors
11	BRS	L32	1	bacteriophage adj "77" adj orf adj "104"	USPAT; US-PGPUB; EPO; JPO; DERWENT	2004/01/16 16:40			0
12	BRS	L39	1	(32 or 34) same inhibitor	USPAT; US-PGPUB; EPO; JPO; DERWENT	2004/01/16 16:40			0
13	BRS	L40	7	dnac adj helicase	USPAT; US-PGPUB; EPO; JPO; DERWENT	2004/01/16 16:40			0
14	BRS	L41	1	31 same 40	USPAT; US-PGPUB; EPO; JPO; DERWENT	2004/01/16 16:42			0
15	BRS	L42	3	31 same (fragment or variant)	USPAT; US-PGPUB; EPO; JPO; DERWENT	2004/01/16 16:43			0
16	BRS	L43	0	36 same 42	USPAT; US-PGPUB; EPO; JPO; DERWENT	2004/01/16 16:43			0
17	BRS	L44	10	pelletier adj jerry.in.	USPAT; US-PGPUB; EPO; JPO; DERWENT	2004/01/16 16:44			0
18	BRS	L46	12	gros adj philippe.in.	USPAT; US-PGPUB; EPO; JPO; DERWENT	2004/01/16 16:44			0
19	BRS	L47	1	dubow adj mike.in.	USPAT; US-PGPUB; EPO; JPO; DERWENT	2004/01/16 16:45			0
20	BRS	L48	1	(44 or 46 or 47) and 31	USPAT; US-PGPUB; EPO; JPO; DERWENT	2004/01/16 16:45			0

**THIS PAGE BLANK (USPTO)**

	Type	L #	Hits	Search Text	DBs	Time Stamp	Comments	Error Definition	Error
1	BRS	L1	24	dnai	USPAT; US-PGPUB; EPO; JPO; DERWENT	2004/01/17 14:58			0
2	BRS	L2	37748	bacteri\$2 same inhibits3	USPAT; US-PGPUB; EPO; JPO; DERWENT	2004/01/17 14:58			0
3	BRS	L3	2	1 same 2	USPAT; US-PGPUB; EPO; JPO; DERWENT	2004/01/17 14:59			0
4	BRS	L4	1	bacteriophage adj "77" adj orf adj "104"	USPAT; US-PGPUB; EPO; JPO; DERWENT	2004/01/17 15:03			0
5	BRS	L5	2	3 same 1	USPAT; US-PGPUB; EPO; JPO; DERWENT	2004/01/17 15:00			0
6	BRS	L6	1	4 same 1	USPAT; US-PGPUB; EPO; JPO; DERWENT	2004/01/17 15:00			0
7	BRS	L7	24094	bacteriophage	USPAT; US-PGPUB; EPO; JPO; DERWENT	2004/01/17 15:01			0
8	BRS	L8	434	2 same 7	USPAT; US-PGPUB; EPO; JPO; DERWENT	2004/01/17 15:01			0
9	BRS	L9	5	bacteriophage adj inhibits3 adj protein	USPAT; US-PGPUB; EPO; JPO; DERWENT	2004/01/17 15:03			0
10	BRS	L10	0	1 same 9	USPAT; US-PGPUB; EPO; JPO; DERWENT	2004/01/17 15:04			0

**THIS PAGE BLANK (USPTO)**

=> d his

(FILE 'HOME' ENTERED AT 16:50:41 ON 16 JAN 2004)

FILE 'MEDLINE, CAPLUS, BIOSIS, EMBASE, SCISEARCH, AGRICOLA'  
ENTERED AT

16:51:01 ON 16 JAN 2004

L1 223 S STAPHYLOCOCCUS AUREUS  
L2 53545 S S. AUREUS  
L3 53722 S L1 OR L2  
L4 148 S DNAI  
L5 3 S L3 (P) L4  
L6 3 S INHIBIT? (P) L5  
L7 2 DUPLICATE REMOVE L6 (1 DUPLICATE REMOVED)  
L8 375307 S BACTERI? (P) INFECTION  
L9 3 S L4 (P) L8  
L10 3 DUPLICATE REMOVE L9 (0 DUPLICATES REMOVED)  
L11 1 S L10 NOT L7  
L12 2 S BACTERIOPHAGE 77 ORF 104  
L13 2 S L12 (P) L4  
L14 0 S L13 NOT L9  
L15 2 S L4 (P) INHIBITOR  
L16 2 S L15 NOT L9  
L17 1 DUPLICATE REMOVE L16 (1 DUPLICATE REMOVED)  
L18 16 S L4 (P) (FRAGMENT OR VARIANT)  
L19 3 S L18 (P) L3  
L20 1 S L19 NOT L9  
L21 0 S PELLETIER J?/AU  
L22 3615 S PELLETIER J?/AU  
L23 2074 S GROS P?/AU  
L24 393 S DUBOW M?/AU  
L25 6047 S (L22 OR L23 OR L24)  
L26 3 S L25 AND L4  
L27 2 DUPLICATE REMOVE L26 (1 DUPLICATE REMOVED)  
L28 0 S L27 NOT L9

=> log y

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 15:11:40 ; Search time 23.912 Seconds

(without alignments)  
1088.625 Million cell updates/sec

Title: US-09-689-952-16

Perfect score: 851  
Sequence: 1 AADICTAITNGEQVKGTL.....ERVKSISTPYFLSGENFRNN 164

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq\_19jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	851	100.0	164	22	AA47315
2	851	100.0	250	22	AA47316
3	851	100.0	313	22	AA47317
4	782	91.9	299	22	AA682049
5	782	91.9	299	22	AA682337
6	782	91.9	307	23	ABP38947
7	418.5	49.2	307	23	AB49355
8	393	46.2	92	23	ABP0981
9	351	41.2	300	23	ABP27829

10	346	40.7	298	24	ABU02187
11	332.5	39.1	300	23	ABP27828
12	330	38.8	293	23	ABP54065
13	104.5	12.3	248	22	AA698965
14	104	12.0	386	22	ABG25088
15	102.5	12.0	304	23	ABP40308
16	101	11.9	294	23	ABP54759
17	100	11.8	291	23	ABP53757
18	99	11.6	855	22	AA695545
19	98.5	11.6	281	22	ABG05583
20	98.5	11.6	281	22	ABG18846
21	98	11.5	467	22	AAU34802
22	98	11.5	479	22	ABG18303
23	97.5	11.5	798	22	AA696715
24	97	11.4	435	22	ABG25176
25	95.5	11.2	340	21	AAV52000
26	95.5	11.2	340	21	AAV51629
27	95.5	11.2	441	19	AAW80725
28	95.5	11.2	442	22	AAU37959
29	95.5	11.2	453	24	ABU02771
30	95.5	11.2	784	22	ABG30180
31	95	11.2	383	24	ABP78794
32	94	11.0	363	22	AA696616
33	94	11.0	451	23	ABP49805
34	94	11.0	453	23	ABP30561
35	94	11.0	484	23	ABP27967
36	93.5	11.0	892	24	ABG74125
37	93	10.9	500	23	ABP65636
38	92.5	10.9	574	22	AAU17209
39	92.5	10.9	893	24	ABG74127
40	92.5	10.9	893	24	ABJ19332
41	92.5	10.9	955	24	ABG74126
42	92	10.8	264	18	AAW21778
43	92	10.8	433	24	ABP77452
44	92	10.8	439	22	ABP61378
45	92	10.8	440	17	AA694600

#### ALIGNMENTS

RESULT 1	AA47315	AA47315 standard; Protein; 164 AA.
XX	AA47315;	
XX	29-AUG-2001 (first entry)	
DE	Amino acids 150-313 of S. aureus DnaI.	
XX	DnaI; S. aureus; inhibitor; bacteriophage 77; ORF 104; phage 77ORF104;	
KW	screening assay.	
XX	Staphylococcus aureus.	
OS	Staphylococcus aureus.	
XX	WO200146383-A2.	
XX	28-JUN-2001.	
PD	21-DEC-2000; 2000WO-US35180.	
PF	22-DEC-1999; 99US-0470512.	
PR	12-OCT-2000; 2000US-0689952.	
XX	(PHAG-) PHAGETECH INC.	
PA	(WILL/) WILLIAMS K M.	
XX	Pelletier J, Gros P, Dubow M;	
XX	WPI; 2001-418052/44.	
DR	N-PSDB; AAC86104.	
XX		

S. pneumoniae type  
Streptococcus poly  
Lactococcus lactis  
E. coli growth and  
Novel human diago  
Staphylococcus epi  
Lactococcus lactis  
Lactococcus lactis  
Putative P. abyssi  
Novel human diago  
Novel human diago  
E. coli cellular p  
Novel human diago  
Novel human diago  
Novel human diago  
Human ACII protein  
Human ACII protein  
S. pneumoniae chro  
Streptococcus pneu  
S. pneumoniae type  
Novel human diago  
N. gonorrhoeae am  
Putative P. abyssi  
Lacteria monocyclo  
Streptococcus poly  
Streptococcus poly  
Mouse spermatogene  
Bifidobacterium lo  
Novel signal trans  
Human spermato gene  
NOVX related prote  
Human spermato gene  
Protein encoded by  
N. gonorrhoeae am  
Drosophila melanog  
S4 protein. Homo

PT Novel DnaI polypeptides useful for treating and diagnosing microbial,  
 PT preferably bacterial, diseases such as those caused by *Staphylococcus*  
 PT aureus -  
 XX  
 XX  
 XX  
 PS Claim 41; Fig 15; 107pp; English.  
 CC This sequence shows a DnaI polypeptide derived from *S. aureus*. *S.*  
 CC aureus DnaI interacts with a growth-inhibitory bacteriophage 77 ORF  
 CC 104 gene product, to form the basis of a screening assay. DnaI  
 CC polypeptides and polynucleotides are useful for treating microbial,  
 CC preferably bacterial, especially *Staphylococcus*, infections. DnaI  
 CC polypeptides and polynucleotides are useful for biological, diagnostic,  
 CC prophylactic, clinical and therapeutic use, and as components in  
 CC databases useful for search analyses as well as in sequence analysis  
 CC algorithms.  
 CC  
 XX  
 XX  
 SQ Sequence 164 AA;  
 Query Match 100.0%; Score 851; DB 22; Length 164;  
 Best Local Similarity 100.0%; Pred. No. 9.6e-87;  
 Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AADICTAITNGEQVGLVYGPFGTGSFTLGAIANQLSKKVRSTIYLPFIRTLKG 60  
 DB 1 AADICTAITNGEQVGLVYGPFGTGSFTLGAIANQLSKKVRSTIYLPFIRTLKG 60  
 QY 61 GFDGSEFKLHVRANILMLDIGAEEVTPWVRDEVIGPLHYRWVHELPTFFSNFD 120  
 DB 61 GFDGSEFKLHVRANILMLDIGAEEVTPWVRDEVIGPLHYRWVHELPTFFSNFD 120  
 QY 121 YSELHHLAMTRDGEKTKAARIIRVKSISTPYFLSGENFRNN 164  
 DB 121 YSELHHLAMTRDGEKTKAARIIRVKSISTPYFLSGENFRNN 164  
 RESULT 2  
 AAB47316  
 ID AAB47316 standard; Protein; 250 AA.  
 AC AAB47316;  
 XX  
 XX  
 DT 29-AUG-2001 (first entry)  
 XX  
 XX  
 DE Amino acids 64-313 of *S. aureus* DnaI.  
 XX  
 XX  
 KM DnaI; *S. aureus*; inhibitor; bacteriophage 77; ORF 104; phage 77ORF104;  
 KW screening assay.  
 XX  
 XX  
 OS *Staphylococcus aureus*.  
 PN WO200146383-A2.  
 XX  
 PD 28-JUN-2001.  
 XX  
 PF 21-DEC-2000; 2000WO-US35180.  
 XX  
 PR 22-DEC-1999; 99US-0470512.  
 PR 12-OCT-2000; 2000US-0689952.  
 XX  
 XX  
 PA (PHAG-) PHAGETECH INC.  
 PA (WILL/) WILLIAMS K M.  
 XX  
 PI Pelletier J, Gros P, Dubow M;  
 XX  
 DR WPI; 2001-418052/44.  
 XX  
 PT Novel DnaI polypeptides useful for treating and diagnosing microbial,  
 PT preferably bacterial, diseases such as those caused by *Staphylococcus*  
 PT aureus -  
 XX  
 XX  
 PS Disclosure; Fig 15; 107pp; English.  
 CC This sequence shows a DnaI polypeptide derived from *S. aureus*. *S.*

CC aureus DnaI interacts with a growth-inhibitory bacteriophage 77 ORF  
 CC 104 gene product, to form the basis of a screening assay. DnaI  
 CC polypeptides and polynucleotides are useful for treating microbial,  
 CC preferably bacterial, especially *Staphylococcus*, infections. DnaI  
 CC polypeptides and polynucleotides are useful for biological, diagnostic,  
 CC prophylactic, clinical and therapeutic use, and as components in  
 CC databases useful for search analyses as well as in sequence analysis  
 CC algorithms.  
 CC  
 XX  
 XX  
 SQ Sequence 250 AA;  
 Query Match 100.0%; Score 851; DB 22; Length 250;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-86;  
 Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AADICTAITNGEQVGLVYGPFGTGSFTLGAIANQLSKKVRSTIYLPFIRTLKG 60  
 DB 87 AADICTAITNGEQVGLVYGPFGTGSFTLGAIANQLSKKVRSTIYLPFIRTLKG 146  
 QY 61 GFDGSEFKLHVRANILMLDIGAEEVTPWVRDEVIGPLHYRWVHELPTFFSNFD 120  
 DB 147 GFDGSEFKLHVRANILMLDIGAEEVTPWVRDEVIGPLHYRWVHELPTFFSNFD 206  
 QY 121 YSELHHLAMTRDGEKTKAARIIRVKSISTPYFLSGENFRNN 164  
 DB 207 YSELHHLAMTRDGEKTKAARIIRVKSISTPYFLSGENFRNN 250  
 RESULT 3  
 AAB47317  
 ID AAB47317 standard; Protein; 313 AA.  
 AC AAB47317;  
 XX  
 XX  
 DT 29-AUG-2001 (first entry)  
 XX  
 XX  
 DE *S. aureus* DnaI.  
 XX  
 XX  
 KM DnaI; *S. aureus*; inhibitor; bacteriophage 77; ORF 104; phage 77ORF104;  
 KW screening assay.  
 XX  
 XX  
 OS *Staphylococcus aureus*.  
 PN WO200146383-A2.  
 XX  
 PD 28-JUN-2001.  
 XX  
 PF 21-DEC-2000; 2000WO-US35180.  
 XX  
 PR 22-DEC-1999; 99US-0470512.  
 PR 12-OCT-2000; 2000US-0689952.  
 XX  
 XX  
 PA (PHAG-) PHAGETECH INC.  
 PA (WILL/) WILLIAMS K M.  
 XX  
 PI Pelletier J, Gros P, Dubow M;  
 XX  
 DR WPI; 2001-418052/44.  
 DR N-PSDB; AAC86105.  
 XX  
 PT Novel DnaI polypeptides useful for treating and diagnosing microbial,  
 PT preferably bacterial, diseases such as those caused by *Staphylococcus*  
 PT aureus -  
 XX  
 XX  
 PS Disclosure; Fig 1; 107pp; English.  
 CC This sequence shows DnaI derived from *S. aureus*. *S. aureus* DnaI  
 CC interacts with a growth-inhibitory bacteriophage 77 ORF 104  
 CC gene product, to form the basis of a screening assay. DnaI  
 CC polypeptides and polynucleotides are useful for treating microbial,  
 CC preferably bacterial, especially *Staphylococcus*, infections. DnaI  
 CC polypeptides and polynucleotides are useful for biological, diagnostic,  
 CC prophylactic, clinical and therapeutic use, and as components in

CC databases useful for search analyses as well as in sequence analysis  
CC algorithms.

XX Sequence 313 AA;

Query Match 100.0%; Score 851; DB 22; Length 313;  
Best Local Similarity 100.0%; Pred. No. 2.3e-86;  
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AADDTCTAATNGEQYKGLYLPFGPGTGSFIIIGALINOLKSKVNSTIYLPDEFIRTLKG 60  
DB 150 AADDTCTAATNGEQYKGLYLPFGPGTGSFIIIGALINOLKSKVNSTIYLPDEFIRTLKG 209  
QY 61 GPKDGSFEKKLRHVRVANIIMDDIGAEVTPWVRDEVIGPLHYRMVHELPTFFSSNFD 120  
DB 210 GPKDGSFEKKLRHVRVANIIMDDIGAEVTPWVRDEVIGPLHYRMVHELPTFFSSNFD 269  
QY 121 YSELEHHLAMTRDGEKTKARIRIERYKSLSTPYPLSGENFRNN 164  
DB 270 YSELEHHLAMTRDGEKTKARIRIERYKSLSTPYPLSGENFRNN 313

## RESULT 4

AA682049  
ID AAG82049 standard; Protein; 299 AA.

AC AAG82049;

DT 03-SEP-2001 (first entry)

DE S. epidermidis open reading frame protein sequence SEQ ID NO:1192.

KM Staphylococcus epidermidis SRI strain; infection; diagnosis;  
KW vaccination; endocarditis.

OS Staphylococcus epidermidis.

PN WO200134809-A2.

PD 17-MAY-2001.

PF 09-NOV-2000; 2000WO-US30782.

PR 09-NOV-1999; 99US-0164258.

PA (GLAX ) GLAXO GROUP LTD.

PI Kimmerly MJ;

DR WPI: 2001-316495/33.

DR N-PSDB; AAH52899.

PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
useful for vaccinating against infections, e.g. endocarditis -  
Claim 18; Page 342; 2188bp; English.

AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
(II), given in AAG81454 to AAG81120, from Staphylococcus epidermidis.  
(I) and (II) can have antibacterial activity and therefore can be used  
in vaccination. The nucleic acids (I) may be used to produce the  
S. epidermidis polypeptides (II) via the production of vectors  
containing them which are used to produce hosts cells which express the  
polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
used to vaccinate subjects and to raise antibodies against the bacteria.  
The polypeptides may also be used to assay for other inhibitors of their  
activity and therefore identify compounds that may be used for the  
treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
AAH5090 represent specifically claimed S. epidermidis genomic DNA  
polynucleotide sequences from the present invention. AAH5091 to  
AAH5098 represent oligonucleotide sequences and primers which are used  
in the exemplification of the present invention.  
N.B. The present invention specifically claims all the polynucleotide

CC sequences given in the sequence listing of the present specification,  
CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
CC no sequences are present for SEQ ID NO:4455 to 4464.

XX Sequence 299 AA;

Query Match 91.9%; Score 782; DB 22; Length 299;  
Best Local Similarity 89.6%; Pred. No. 1.1e-78;  
Matches 147; Conservative 12; Mismatches 5; Indels 0; Gaps 0;

QY 1 AADDTCTAATNGEQYKGLYLPFGPGTGSFIIIGALINOLKSKVNSTIYLPDEFIRTLKG 60  
DB 136 AADDTCTAATNGEQYKGLYLPFGPGTGSFIIIGALINOLKSKVNSTIYLPDEFIRTLKG 195  
QY 61 GPKDGSFEKKLRHVRVANIIMDDIGAEVTPWVRDEVIGPLHYRMVHELPTFFSSNFD 120  
DB 196 GPKDGSFEKKLRHVRVANIIMDDIGAEVTPWVRDEVIGPLHYRMVHELPTFFSSNFD 255  
QY 121 YSELEHHLAMTRDGEKTKARIRIERYKSLSTPYPLSGENFRNN 164  
DB 256 YSELEHHLAMTRDGEKTKARIRIERYKSLSTPYPLSGENFRNN 299

## RESULT 5

AA682337  
ID AAG82337 standard; Protein; 299 AA.

AC AAG82337;

DT 03-SEP-2001 (first entry)

DE S. epidermidis open reading frame protein sequence SEQ ID NO:11768.

KM Staphylococcus epidermidis SRI strain; infection; diagnosis;  
KW vaccination; endocarditis.

OS Staphylococcus epidermidis.

PN WO200134809-A2.

PD 17-MAY-2001.

PF 09-NOV-2000; 2000WO-US30782.

PR 09-NOV-1999; 99US-0164258.

PA (GLAX ) GLAXO GROUP LTD.

PI Kimmerly MJ;

DR WPI: 2001-316495/33.

DR N-PSDB; AAH53187.

PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
useful for vaccinating against infections, e.g. endocarditis -  
Claim 18; Page 488; 2188bp; English.

AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
(II), given in AAG81454 to AAG81120, from Staphylococcus epidermidis.  
(I) and (II) can have antibacterial activity and therefore can be used  
in vaccination. The nucleic acids (I) may be used to produce the  
S. epidermidis polypeptides (II) via the production of vectors  
containing them which are used to produce hosts cells which express the  
polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
used to vaccinate subjects and to raise antibodies against the bacteria.  
The polypeptides may also be used to assay for other inhibitors of their  
activity and therefore identify compounds that may be used for the  
treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
AAH5090 represent specifically claimed S. epidermidis genomic DNA  
polynucleotide sequences from the present invention. AAH5091 to  
AAH5098 represent oligonucleotide sequences and primers which are used

CC In the exemplification of the present invention.  
CC N.B. The present invention specifically claims all the polynucleotide  
CC sequences given in the sequence listing of the present specification,  
CC however the sequence listing only goes up to SEQ ID NO:4455 so even  
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
CC no sequences are present for SEQ ID NO:4455 to 4464.

XX SQ Sequence 299 AA;

Query Match 91.9%; Score 782; DB 22; Length 299;

Best Local Similarity 89.6%; Pred. No. 1,1e-78;  
Matches 147; Conservative 12; Mismatches 5; Indels 0; Gaps 0;

QY 1 AADDICTAATNGEQVGLYLYGPGTGKSPFIIGALANQKSKKRSSTIYLPFFIRTLKG 60  
DB 136 AADQICTAITNDKVKGLYLYGPGTGKSPFIIGALANQKSKKRSSTIYLPFFIRTLKG 195  
QY 61 GFDGSGFEKKLHRVREANILMLDIDGAEVTPWVRDEVIGPLHYRMVHELPTFFSSNPD 120  
DB 196 GFDGSGFEKKLQVRREANILMLDIDGAEVTPWVRDEVIGPLHYRMVHELPTFFSSNPN 255  
QY 121 YSELHHLMTNRDGEKTKARIIERVKSLSTPYPLSGENFRNN 164  
DB 256 YSELHHLSTTRDGTERTKARIIERIKITLSTPYPLTGKNFRNN 299

#### RESULT 6

ABP38947  
ID ABP38947 standard; Protein; 307 AA.

XX ABP38947;

XX 24-JUL-2002 (first entry)

DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3792.

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
KM antibacterial; gene therapy.

XX Staphylococcus epidermidis.

XX US6380370-B1.

XX 30-APR-2002.

XX 13-AUG-1998; 98US-0134001.

XX 14-AUG-1997; 97US-055779P.

XX 08-NOV-1997; 97US-064964P.

XX (GENO-) GENOME THERAPEUTICS CORP.

PI Doucette-Stamm LA, Bush D;

DR WPI; 2002-381255/41.

DR N-PSDB; ABN91492.

PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
PT polypeptide, useful for diagnosing and treating bacterial infections -

XX Disclosure; SEQ ID 3792; 267pp; English.

CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
CC antibacterial activity and can be used in gene therapy. The sequences  
CC can also be used in the diagnosis and treatment of bacterial infections,  
CC particularly S. epidermidis infections. The sequences can be used to  
CC screen for compounds able to interfere with the S. epidermidis life  
CC cycle or inhibit S. epidermidis infection.  
CC N.B.: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC USPTO web site.

XX SQ Sequence 307 AA;

Query Match 91.9%; Score 782; DB 23; Length 307;

Best Local Similarity 89.6%; Pred. No. 1,1e-78;  
Matches 147; Conservative 12; Mismatches 5; Indels 0; Gaps 0;

QY 1 AADDICTAATNGEQVGLYLYGPGTGKSPFIIGALANQKSKKRSSTIYLPFFIRTLKG 60  
DB 144 AADQICTAITNDKVKGLYLYGPGTGKSPFIIGALANQKSKKRSSTIYLPFFIRTLKG 203  
QY 61 GFDGSGFEKKLHRVREANILMLDIDGAEVTPWVRDEVIGPLHYRMVHELPTFFSSNPD 120  
DB 204 GFDGSGFEKKLQVRREANILMLDIDGAEVTPWVRDEVIGPLHYRMVHELPTFFSSNPN 263  
QY 121 YSELHHLMTNRDGEKTKARIIERVKSLSTPYPLSGENFRNN 164  
DB 264 YSELHHLSTTRDGTERTKARIIERIKITLSTPYPLTGKNFRNN 307

#### RESULT 7

ABB49355  
ID ABB49355 standard; Protein; 307 AA.

XX ABB49355;

XX 05-FEB-2002 (first entry)

DE Listeria monocytogenes protein #2059.

XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
KM vitamin B12; bacterial infection; disease.

XX Listeria monocytogenes.

XX WO200177335-A2.

XX 18-OCT-2001.

XX 11-APR-2001; 2001WO-FR01118.

XX 11-APR-2000; 2000FR-0004629.

XX (INSP ) INST PASTEUR.

PI Buchrieser C, Frangeul L, Couve E, Ruaniok C, Feihl H, Dehoux P;

PI Dusserre O, Chetouani F, Nedjari H, Glaeser P, Kuntz F, Cosserat P;

PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;

PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;

PI Chakraborty T, Demann B, Hain T, Berche P, Chardot A, Durant L;

PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;

PI Madueno E, De Pablo B, Weiland J, Kaerst U, Entian K, Hauf J;

PI Rose M, Voss H;

DR WPI; 2002-010914/01.

PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
PT and prevention of Listeria and related bacterial infections, and  
PT related polypeptides -

XX Claim 6; SEQ ID NO 2060; 192pp; French.

CC The present invention relates to the genome sequence of Listeria  
CC monocytogenes EGD-e (see ABN03041). The genome sequence and fragments of  
CC it are useful for selecting probes and primers for detecting genes in L.  
CC monocytogenes and related organisms, and for studying genetic  
CC polymorphisms and other genomes. The present sequence is a protein  
CC encoded by the genome sequence of the present invention. Proteins  
CC expressed from the genome sequence are useful for raising specific  
CC antibodies, identification of L. monocytogenes and related organisms, and  
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
CC B12. The genome sequence and proteins encoded by it are also useful for  
CC selecting compounds that regulate gene expression and cell replication

CC The present invention describes substantially purified human proteins  
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
CC in the specification). ABN15762 to ABN27252 encode the human ORFX

DR N-PSDB; ABN68460.

PT New Streptococcus protein for the treatment or prevention of infection  
PT or disease caused by Streptococcus bacteria, such as meningitis, and  
PT for detecting a compound that binds to the protein -

PS Claim 1; Page 3645; 4525pp; English.

CC The invention relates to a protein (ABP25413-ABP30895) from group B  
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
CC the specification. The proteins have antibacterial and antiinflammatory  
CC activity. (1), nucleic acids encoding (1), ABN66044-ABN7156 and  
CC antibodies that bind (1) are used in the manufacture of medicaments for  
CC the treatment or prevention of infection or disease caused by  
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
CC Nucleic acids encoding (1) are used to detect Streptococcus in a  
CC biological sample. (1) is used to determine whether a compound binds to  
CC (1). A composition comprising (1) or a nucleic acid encoding (1), may be  
CC used as a vaccine or diagnostic composition. The disease caused by  
CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
CC acid encoding (1) may be used to recombinantly produce (1) and may be  
CC used in gene therapy. Antibodies to (1) are used for affinity  
CC chromatography, immunoassays, and distinguishing/identifying  
CC Streptococcus proteins.

XX Sequence 300 AA;

Query Match 41.2%; Score 351; DB 23; Length 300;

Best Local Similarity 43.1%; Pred. No. 1,3e-30; Mismatches 53; Indels 2; Gaps 2;

Matches 66; Conservative 36; Mismatches 53; Indels 2; Gaps 2;

4 DICTAITNGEQVKGVLVYGGPGTGKSPFLGAIANOLSKK-VNSTIITYLDEPFRITLKGCF 62

142 DPEQVPSAEQ-KGLVYGDVGIGKSYLLAMAHLESEKGVSTTLHPSPFAIDVGNAL 200

63 KDSFEKKHRYVREANILMDIGAEVTPWVRDEVIGPLIHYRMVHELTPFSSNDYS 122

201 SNCSVKEIDAVKVPFLILDIGAEQATSWRDEVQVLLQYRMLELPTFTSNYSFA 260

123 ELEHNLMTDRGEKTKYARIIRVKSISTPYFLSGENFR 162

261 DLRKWTATIKGSDPTWQAKVMERYRLAEPHLEGANRR 300

RESULT 10

ABU02187

ID ABU02187 standard; Protein; 298 AA.

XX ABU02187;

DT 11-FEB-2003 (first entry)

DE S. pneumoniae type 4 strain protein from coding region #1764.

XX Bacterial meningitis; pneumonia; sepsis; otitis media;

KW ear infection; antiinflammatory; antibacterial; immunostimulant;

XX auditory; respiratory; gene therapy; vaccine.

OS Streptococcus pneumoniae type 4 strain.

XX MO200277021-A2.

PD 03-OCT-2002.

XX 27-MAR-2002; 2002WO-IB02163.

XX 27-MAR-2001; 2001GB-0007658.

XX (CHIR-) CHIRON SPA.

XX (GENO-) INST GENOMIC RES.

XX Masignani V, Tettein H, Frazer C;

XX WPI; 2003-040579/03.

DR N-PSDB; ABX07476.

XX

XX

PT

PT

PT

PT

PT

PT

PT

PT

PT

PT

PT

PT

PT

PT

PT

PT

PT

PT

PT

PT

PT

PT

PT

PT

PT

PT

PT

PT

PT

PT

PT

PT

PT

PT

PT

PT

New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
useful as medicaments for treating or preventing a disease or infection  
due to streptococcus bacteria, such as pneumonia, sepsis, otitis media  
or ear infection -

PS Claim 1; SEQ ID NO 3528; 56pp; English.

CC The invention relates to a protein comprising or having at least 50%  
CC identity to any of the 2469 amino acid sequences identified in the  
CC specification (available on a computer readable format), or its fragment,  
CC expressed from 2469 of 2489 identified DNA coding regions from the  
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
CC AB556454. Also included are an antibody which binds one of the  
CC proteins, treating a patient by administering the protein, DNA or  
CC antibody (in a composition), a kit comprising first and second primers,  
CC which are the nucleic acid cited above or fragments between nucleotides  
CC 8-100 of a sequence not defined in the specification, for amplifying a  
CC target sequence contained within a Streptococcus nucleic acid sequence,  
CC where the first primer is substantially complementary to the target  
CC sequence and the second primer is substantially complementary to the  
CC complement of the target sequence, and where the parts of the primers  
CC having substantial complementarity define the termini of the target  
CC sequence to be amplified, assay comprising contacting a test compound  
CC with the protein, and determining whether the test compound binds to the  
CC protein and a Streptococcus pneumoniae bacterium, where one or more  
CC genes encoding the proteins has been rendered inactive. The proteins,  
CC nucleic acid molecules, antibody and compositions are useful as  
CC medicaments for treating or preventing a disease or infection due to  
CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,  
CC sepsis, otitis media or ear infection. They are also useful in developing  
CC vaccines, diagnostics and antibiotics. The methods are useful for  
CC identifying immunodominant proteins. The present sequence is one of  
CC the 2469 proteins expressed by the identified coding regions from the  
CC genomic sequence.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 298 AA;

Query Match 40.7%; Score 346; DB 24; Length 298;

Best Local Similarity 44.3%; Pred. No. 4.8e-30; Mismatches 46; Indels 2; Gaps 2;

Matches 66; Conservative 35; Mismatches 46; Indels 2; Gaps 2;

QY 16 KGLVYGGPGTGKSPFLGAIANOLSKK-VNSTIITYLDEPFRITLKGCFSGFEKKHRYV 74

DB 150 KGLVYGDVGIGKSYLLAMAHLESEKGVSTTLHPSPFAIDVGNALSDGVNKTLYDRI 209

QY 75 REANILMDIGAEVTPWVRDEVIGPLIHYRMVHELTPFSSNDYSLEHNLMTDRG 134

DB 210 KJSEVILLIDIGAEQATSWRDEVQVLLQYRMLELPTFTSNYSFADLEKHPAYKKG 269

QY 135 EKKTKYARIIRVKSISTPYFLSGENFR 162

DB 270 NDETWARVRMERIRYLAETRLLEGVNR 298

RESULT 11

ABP27828

ID ABP27828 standard; Protein; 300 AA.

XX ABP27828;

DT 02-JUL-2002 (first entry)

DE Streptococcus polypeptide SEQ ID NO 4832.

XX Streptococcus; GAS; group B streptococcus; Streptococcus agalactiae;

XX group A streptococcus; Streptococcus pyogenes; antibacterial;

XX antiinflammatory; infection; vaccine; meningitis; gene therapy.

[illegible][illegible]

PD 17-MAY-2001.  
 XX 09-NOV-2000; 2000MO-US0950.  
 XX 09-NOV-1999; 99US-0164415.  
 XX (ELIT-) ELITRA PHARM INC.  
 XX Foreyth RA, Ohlsen K, Zyskind J,  
 XX WPI; 2001-335933/35.  
 DR N-PSDB; AAH84636.  
 PT Novel nucleic acids that inhibit Escherichia coli proliferation, useful  
 PT for screening for homologous genes and for designing expression vectors  
 PS  
 XX Claim 19; Page 492; 522pp; English.  
 CC AAH84373 to AAH84499 represent Escherichia coli growth and proliferation  
 CC related DNA sequences (I). AAH84500 to AAH84670 encode the E. coli  
 CC growth and proliferation related proteins given in AAG99078 and AAG98830  
 CC to AAG98999. (I) can be used as potential targets for the generation of  
 CC new antimicrobial agents, and for identification of compounds which  
 CC interact with the gene products of (I). In addition the expression of  
 CC (I) and the purification of the proteins, the purified proteins can be  
 CC used to generate reagents and screen small molecule libraries or other  
 CC candidate compound libraries for compounds that can be further developed  
 CC to yield novel antimicrobial compounds. In addition, nucleic acid probes  
 CC complementary to (I) that are specific for particular species of  
 CC microorganisms can be used to identify particular microorganism species  
 CC in clinical specimens, therefore, providing a rapid and dependable  
 CC method by which to identify the causative agents of a bacterial  
 CC infection. Also, antibodies generated against proteins translated from  
 CC mRNA transcribed from proliferation-related sequences can also be used  
 CC to screen for specific microorganisms that produce such proteins in a  
 CC species-specific manner. AAH84371 and AAH84670 represent sequencing  
 CC primers used in the isolation of E. coli growth and proliferation  
 CC related sequence, which are used in an example from the present  
 CC invention.  
 XX  
 XX Sequence 248 AA;  
 SQ  
 Query Match 12.3%; Score 104.5; DB 22; Length 248;  
 Best Local Similarity 27.2%; Pred. No. 0.0033;  
 Matches 40; Conservative 28; Mismatches 60; Indels 19; Gaps 6;  
 QY 2 ADDICATITNGEYVKGILYSPFGKSFILAIANOLSKKXKVRSTIYLPPIRTLXKG 61  
 DB 94 ADELMTGCTN-----FASGKPGTGNHIAAIGNLLDKDGTIVIVTVADVMSALHAS 147  
 QY 62 FKDG-SFEKKLHVRVREANILMDIGAEEVTPWVRDEVIGPLHYRWVHELPTFFSSNPD 120  
 DB 148 YDDGSGEKFRLCELDVLVDDEIGIQRFT---KNEQV--VLH-QIVDRRTASMSVGM 201  
 QY 121 YSELEHHILAMTRDGEERTKAARIIRY 147  
 DB 202 LTNINYEAMKTLIGE-----RIMDRM 222  
 RESULT 14  
 ABG25088  
 ID ABG25088 standard; Protein; 386 AA.  
 XX  
 AC ABG25088;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 XX Novel human diagnostic protein #25079.  
 DE  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KM food supplement; medical imaging; diagnostic; genetic disorder.  
 XX

OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 XX 11-OCT-2001.  
 PD  
 XX 30-MAR-2001; 2001MO-US08631.  
 PF  
 XX 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 PA  
 XX Drmanac RT, Liu C, Tang YT,  
 XX WPI; 2001-639362/73.  
 DR N-PSDB; AAS89275.  
 PT  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 PS  
 XX Claim 20; SEQ ID No 55447; 103pp; English.  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of. (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 386 AA;  
 SQ  
 Query Match 12.2%; Score 104; DB 22; Length 386;  
 Best Local Similarity 26.4%; Pred. No. 0.0068;  
 Matches 43; Conservative 33; Mismatches 67; Indels 20; Gaps 7;  
 QY 2 ADDICATITNGEYVKGILYSPFGKSFILAIANOLSKKXKVRSTIYLPPIRTLXKG 61  
 DB 188 ADELMTGCTN-----FASGKPGTGNHIAAIGNLLDKDGTIVIVTVADVMSALHAS 241  
 QY 62 FKDG-SFEKKLHVRVREANILMDIGAEEVTPWVRDEVIGPLHYRWVHELPTFFSSNPD 120  
 DB 242 YDDGSGEKFRLCELDVLVDDEIGIQRFT---KNEQV--VLH-QIVDRRTASMSVGM 295  
 QY 121 YSELEHHILAMTRDGEERTKAARIIRVYSLSTPYPLSGENFEN 163  
 DB 296 LTNINYEAMKTLIGE-----RIMDRM-TMNGRWATHNNDN 331  
 RESULT 15  
 ABP40308  
 ID ABP40308 standard; Protein; 304 AA.  
 XX  
 AC ABP40308;  
 XX  
 DT 24-JUL-2002 (first entry)  
 XX

XX	Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5153.
DE	
XX	Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW	antibacterial; gene therapy.
XX	
OS	Staphylococcus epidermidis.
XX	
FN	US6380370-B1.
XX	
PD	30-APR-2002.
XX	
PF	13-AUG-1998; 98US-0134001.
XX	
PR	14-AUG-1997; 97US-055779P.
PR	08-NOV-1997; 97US-064964P.
XX	
PA	(GENO-) GENOME THERAPEUTICS CORP.
XX	
P1	Doucette-Stamm LA, Bush D;
DR	WPI; 2002-381255/41.
DR	N-PSDB; ABN92853.
XX	
PT	Novel isolated nucleic acid encoding a Staphylococcus epidermidis
XX	polypeptide, useful for diagnosing and treating bacterial infections -
PS	Disclosure; SEQ ID 5153; 267bp; English.
XX	
AB	ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC	frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC	given in ABP95124 to ABP37960. The S. epidermidis sequences have
CC	antibacterial activity and can be used in gene therapy. The sequences
CC	can also be used in the diagnosis and treatment of bacterial infections,
CC	particularly S. epidermidis infections. The sequences can be used to
CC	screen for compounds able to interfere with the S. epidermidis life
CC	cycle or inhibit S. epidermidis infection.
CC	N.B. The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from the
CC	USPTO web site.

**SQ Sequence 304 AA;**

Query Match	12.0%	Score 102.5;	DB 23;	Length 304;
Best Local Similarity	21.4%;	Pred No.0.0073;		
Matches 33;	Conservative 35;	Mismatches 59;	Indels 27;	Gaps 4;

[illegible]

Search completed: December 15, 2003, 15:13:43  
Job time : 24.912 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 15:11:41 ; Search time 8.57221 Seconds  
(without alignments)  
809,474 Million cell updates/sec

Title: US-09-689-952-16

Perfect score: 851  
Sequence: 1 AADDICTAITNGEQVKGSLV.....ERKSLSTPYFLSGENFRNN 164

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	851	100.0	313	4	US-09-470-512A-2	Sequence 2, Appl1
2	849	99.8	313	4	US-09-470-512A-10	Sequence 10, Appl1
3	782	91.9	307	4	US-09-134-001C-3792	Sequence 3792, Ap
4	427.5	50.2	280	4	US-09-470-512A-14	Sequence 14, Appl1
5	420.5	49.4	278	4	US-09-470-512A-15	Sequence 15, Appl1
6	343.5	40.4	310	4	US-09-107-532A-5580	Sequence 5580, Ap
7	105.5	12.4	268	4	US-09-328-352-5963	Sequence 5963, Ap
8	102.5	12.0	304	4	US-09-134-001C-5153	Sequence 5153, Ap
9	90	10.6	389	2	US-08-820-170A-13	Sequence 13, Appl1
10	90	10.6	389	3	US-09-055-699-13	Sequence 13, Appl1
11	90	10.6	389	3	US-09-273-565-13	Sequence 13, Appl1
12	90	10.6	389	4	US-09-565-538-13	Sequence 13, Appl1
13	90	10.6	389	4	US-09-661-468-13	Sequence 13, Appl1
14	90	10.6	389	4	US-09-976-165-13	Sequence 13, Appl1
15	86.5	10.2	1349	2	US-08-612-734B-2	Sequence 2, Appl1
16	86	10.1	529	4	US-09-252-991A-18630	Sequence 18630, A
17	86	10.1	1334	2	US-08-596-545-2	Sequence 2, Appl1
18	86	10.1	1334	3	US-09-328-320-2	Sequence 2, Appl1
19	84	9.9	467	4	US-09-198-452A-443	Sequence 443, Appl1
20	83.5	9.8	456	3	US-09-268-364-21	Sequence 21, Appl1
21	83	9.8	308	4	US-09-149-476-607	Sequence 607, Appl1
22	83	9.8	399	4	US-09-149-476-474	Sequence 474, Appl1
23	83	9.8	806	3	US-08-999-774A-6	Sequence 6, Appl1
24	82	9.6	227	4	US-09-291-170A-9	Sequence 9, Appl1
25	81.5	9.6	266	4	US-09-724-884-9	Sequence 9, Appl1
26	81.5	9.6	266	4	US-09-218-363-8	Sequence 8, Appl1
27	81.5	9.6	267	4	US-09-218-363-10	Sequence 10, Appl1

28	81.5	9.6	290	4	US-09-218-363-4	Sequence 4, Appl1
29	81	9.5	220	4	US-09-179-179-4	Sequence 4, Appl1
30	80.5	9.5	264	4	US-09-328-352-7647	Sequence 7647, Ap
31	80.5	9.5	957	4	US-09-107-532A-4021	Sequence 4021, Ap
32	80	9.4	850	4	US-09-107-532A-4631	Sequence 4631, Ap
33	79.5	9.3	194	4	US-09-218-363-16	Sequence 16, Appl1
34	79.5	9.3	134	4	US-09-218-363-17	Sequence 17, Appl1
35	79.5	9.3	332	4	US-09-218-363-2	Sequence 2, Appl1
36	79.5	9.3	332	4	US-09-218-363-6	Sequence 6, Appl1
37	79.5	9.3	332	4	US-09-218-363-9	Sequence 9, Appl1
38	79.5	9.3	332	4	US-09-149-476-467	Sequence 467, Appl1
39	79.5	9.3	859	4	US-09-199-637A-281	Sequence 281, Appl1
40	79.5	9.3	859	4	US-09-252-991A-21413	Sequence 21413, A
41	78.5	9.2	320	4	US-09-252-991A-32024	Sequence 32024, A
42	78	9.2	215	4	US-09-291-170A-6	Sequence 6, Appl1
43	78	9.2	215	4	US-09-724-884-6	Sequence 6, Appl1
44	78	9.2	226	3	US-09-268-364-2	Sequence 2, Appl1
45	78	9.2	405	2	US-08-222-719-2	Sequence 2, Appl1

#### ALIGNMENTS

```

RESULT 1
US-09-470-512A-2
; Sequence 2, Application US/09470512A
; Patent No. 6376652
; GENERAL INFORMATION:
; APPLICANT: PhageTech, Inc.
; TITLE OF INVENTION: Compositions and methods involving an essential Staphylococcus Au;
; TITLE OF INVENTION: gene and its encoded protein
; FILE REFERENCE: 21715/1000
; CURRENT APPLICATION NUMBER: US/09/470,512A
; CURRENT FILING DATE: 1999-12-12
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 313
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-470-512A-2

Query Match      100.0%; Score 851; DB 4; Length 313;
Best Local Similarity 100.0%; Pred. No. 2.1e-96;
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AADDICTAITNGEQVKGSLVGPFGTGSFLIGALANOLSKKVSSTIYLPEFRTLKG 60
DB      150 AADDICTAITNGEQVKGSLVGPFGTGSFLIGALANOLSKKVSSTIYLPEFRTLKG 209
QY      61 GKQDSFEKCLRVRBANILMDITGAEVTPWVDEVIQPLIHYRMVHELPTFFSSNFD 120
DB      210 GKQDSFEKCLRVRBANILMDITGAEVTPWVDEVIQPLIHYRMVHELPTFFSSNFD 269
QY      121 YSELEHILAMTRDGEKTKAARIIRVKSISTPYFLSGENFRNN 164
DB      270 YSELEHILAMTRDGEKTKAARIIRVKSISTPYFLSGENFRNN 313

RESULT 2
US-09-470-512A-10
; Sequence 10, Application US/09470512A
; Patent No. 6376652
; GENERAL INFORMATION:
; APPLICANT: PhageTech, Inc.
; TITLE OF INVENTION: Compositions and methods involving an essential Staphylococcus Au;
; TITLE OF INVENTION: gene and its encoded protein
; FILE REFERENCE: 21715/1000
; CURRENT APPLICATION NUMBER: US/09/470,512A
; CURRENT FILING DATE: 1999-12-12
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10

```

LENGTH: 313  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-470-512A-10

Query Match 99.8%; Score 849; DB 4; Length 313;  
Best Local Similarity 99.4%; Pred. No. 3.8e-98;  
Matches 163; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AADICTAINTGEOVKGLYVGPFGTGSFLIGAIANQLSKKVRSTIYLPFIRTLKG 60  
DB 150 AADICTAINTGEOVKGLYVGPFGTGSFLIGAIANQLSKKVRSTIYLPFIRTLKG 209  
QY 61 GFDGSEFKLHRYREANIIMLDDIGAEVTPWVRDEVIGPLHYRWVHELPTFFSSNFD 120  
DB 210 GFDGSEFKLHRYREANIIMLDDIGAEVTPWVRDEVIGPLHYRWVHELPTFFSSNFD 269  
QY 121 YSELEHNLAMTRDGEETKARIIERYKSLSTPYFLSGENFRNN 164  
DB 270 YSELEHNLAMTRDGEETKARIIERYKSLSTPYFLSGENFRNN 313

RESULT 3  
US-09-134-001C-3792  
Sequence 3792, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
FILE REFERENCE: GFC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 3792  
LENGTH: 307  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3792

Query Match 91.9%; Score 782; DB 4; Length 307;  
Best Local Similarity 89.6%; Pred. No. 9.5e-90;  
Matches 147; Conservative 12; Mismatches 5; Indels 0; Gaps 0;

QY 1 AADICTAINTGEOVKGLYVGPFGTGSFLIGAIANQLSKKVRSTIYLPFIRTLKG 60  
DB 144 AADICTAINTGEOVKGLYVGPFGTGSFLIGAIANQLSKKVRSTIYLPFIRTLKG 203  
QY 61 GFDGSEFKLHRYREANIIMLDDIGAEVTPWVRDEVIGPLHYRWVHELPTFFSSNFD 120  
DB 204 GFDGSEFKLHRYREANIIMLDDIGAEVTPWVRDEVIGPLHYRWVHELPTFFSSNFD 263  
QY 121 YSELEHNLAMTRDGEETKARIIERYKSLSTPYFLSGENFRNN 164  
DB 264 YSELEHNLAMTRDGEETKARIIERYKSLSTPYFLSGENFRNN 307

RESULT 4  
US-09-470-512A-14  
Sequence 14, Application US/09470512A  
Patent No. 6376652  
GENERAL INFORMATION:  
APPLICANT: PhageTech, Inc.  
TITLE OF INVENTION: Compositions and methods involving an essential Staphylococcus Au  
FILE REFERENCE: 21715/1000  
CURRENT APPLICATION NUMBER: US/09/470,512A  
CURRENT FILING DATE: 1999-12-12  
NUMBER OF SEQ ID NOS: 16

SOFTWARE: Patentin version 3.0  
SEQ ID NO 14  
LENGTH: 280  
TYPE: PRT  
ORGANISM: Bacillus subtilis  
US-09-470-512A-14

Query Match 50.2%; Score 427.5; DB 4; Length 280;  
Best Local Similarity 55.7%; Pred. No. 2.7e-45;  
Matches 83; Conservative 26; Mismatches 39; Indels 1; Gaps 1;

QY 16 KGLYVGPFGTGSFLIGAIANQLSKKVRSTIYLPFIRTLKGFDGSEFKLHRYR 75  
DB 132 KGLYVGPFGTGSFLIGAIANQLSKKVRSTIYLPFIRTLKGFDGSEFKLHRYR 191  
QY 76 EANIIMLDDIGAEVTPWVRDEVIGPLHYRWVHELPTFFSSNFDYSELEHNLAMTRDGE 135  
DB 192 TPEVLMDDIGAEVTPWVRDEVIGPLHYRWVHELPTFFSSNFDYSELEHNLAMTRDGE 251  
QY 136 -EKTARIIERYKSLSTPYFLSGENFRN 163  
DB 252 KEEVKARLIERLILYLAAPRILDGENRRH 280

RESULT 5  
US-09-470-512A-15  
Sequence 15, Application US/09470512A  
Patent No. 6376652  
GENERAL INFORMATION:  
APPLICANT: PhageTech, Inc.  
TITLE OF INVENTION: Compositions and methods involving an essential Staphylococcus Au  
FILE REFERENCE: 21715/1000  
CURRENT APPLICATION NUMBER: US/09/470,512A  
CURRENT FILING DATE: 1999-12-12  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 15  
LENGTH: 278  
TYPE: PRT  
ORGANISM: Bacillus halodurans  
US-09-470-512A-15

Query Match 49.4%; Score 420.5; DB 4; Length 278;  
Best Local Similarity 50.0%; Pred. No. 2e-44;  
Matches 77; Conservative 33; Mismatches 43; Indels 1; Gaps 1;

QY 12 GEOVKGLYVGPFGTGSFLIGAIANQLSKKVRSTIYLPFIRTLKGFDGSEFKL 71  
DB 125 GEOVKGLYVGPFGTGSFLIGAIANQLSKKVRSTIYLPFIRTLKGFDGSEFKL 184  
QY 72 HRYREANIIMLDDIGAEVTPWVRDEVIGPLHYRWVHELPTFFSSNFDYSELEHNLAMTR 131  
DB 185 DFKNAQVLIFFDIDGAEVTPWVRDEVIGPLHYRWVHELPTFFSSNFDYSELEHNLAMTR 244  
QY 132 -RDGEETKARIIERYKSLSTPYFLSGENFRNN 164  
DB 245 DKSGETLAKAKRVERIRHRTVSVWVGQVRREH 278

RESULT 6  
US-09-107-532A-5580  
Sequence 5580, Application US/09107532A  
Patent No. 6585275  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSER: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham

STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 5580:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 310 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (B) LOCATION 1...310  
SEQUENCE DESCRIPTION: SEQ ID NO: 5580:  
US-09-107-532A-5580  
Query Match 40.4%; Score 343.5; DB 4; Length 310;  
Best Local Similarity 46.3%; Pred. No. 1.1e-34;  
Matches 69; Conservative 28; Mismatches 51; Indels 1; Gaps 1;  
QY 16 KELYVYGPFGTSGFLGAIANOLSKKVRSTIYLPFIRTLKGGFKDSEKKLHYR 75  
DB 161 KELYVSGIGIKYTLGATARELAAGTTLVHPTPAVENKQAIQGOVAKLDAYK 220  
QY 76 EANIIMLDIGAEVYTPWVRDEVIGPLHYRWVHELPTFFSSNFDYSELEHILAMTRDG- 134  
DB 221 RPIIMIDIDIGADAMSWIRDEVFVTLQYRMQEQLPFTSSNFTMDLEQHLSTVQRGD 280  
QY 135 EKTQARIIEVKSLSTPYLSENFEN 163  
DB 281 EEPKAKRIMERIRYLTKEIEMTGRNRN 309  
RESULT 7  
US-09-328-352-5963  
Sequence 5963, Application US/09328352  
Patent No. 6562958  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
FILE REFERENCE: GTC99-03PA  
CURRENT APPLICATION NUMBER: US/09/328,352  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 5963  
LENGTH: 268  
TYPE: PRT  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-5963

Query Match 12.4%; Score 105.5; DB 4; Length 268;  
Best Local Similarity 27.0%; Pred. No. 6.9e-05;  
Matches 24; Conservative 19; Mismatches 41; Indels 5; Gaps 2;  
QY 2 ADDICTAI---TNGQVKGVLVYGPFGTSGFLGAIANOLSKKVRSTIYLP--PEFIR 56  
DB 143 AAEFCRKYLTOLGASQHNPLFLYGPFGTGLKTHLMQAVGNALLOAQAPNARVMYMTSQSFVQ 202  
QY 57 TLKGGFKDSEKKLHYRREANIIMLDI 85  
DB 203 DFVSSLQKQKVEEFKRCNSLDLLAVDDI 231  
RESULT 8  
US-09-134-001C-5153  
Sequence 5153, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 5153  
LENGTH: 304  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5153  
Query Match 12.0%; Score 102.5; DB 4; Length 304;  
Best Local Similarity 21.4%; Pred. No. 0.0002;  
Matches 33; Conservative 35; Mismatches 59; Indels 27; Gaps 4;  
QY 2 ADDICTAITNGQVKGVLVYGPFGTSGFLGAIANOLSKKVRSTIYLP-----EPI 55  
DB 135 AASLAVAPAPAAVYPLFLYGVGGLKTHLMALIGHVLSNPNAKVIYTSSEKFTNEFI 194  
QY 56 RTLKGGFKDSEKKLHYRREANIIMLDIGAEVYTPWVRDEVIGPLHYRWVHELPTFF 115  
DB 195 KSI---BDNETAFEREYKIDVLLIDIDIOFTQKQTOEE-----FF 234  
QY 116 SSNFDYSELEHILAMTRDGEKTKARIIEVKYS 149  
DB 235 HTFNEIHONNKQIYISSDRPK-EIAKLEDRLRS 267  
RESULT 9  
US-08-820-170A-13  
Sequence 13, Application US/08820170A  
Patent No. 5831058  
GENERAL INFORMATION:  
APPLICANT: Teutomu, FUMIWARA  
APPLICANT: Takeshi, WATANABE  
APPLICANT: Masato, HORIE  
APPLICANT: Toyomasa, KATAGIRI  
TITLE OF INVENTION: HUMAN GENE  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sughrue, Mion, Zimm, Macpeak & Seas  
STREET: 2100 Pennsylvania Avenue, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: United States  
ZIP: 20037-3202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/820,170A  
 FILING DATE:  
 CLASSIFICATION: 536  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 293-7060  
 TELEFAX: (202) 293-7860  
 TELEX: 6491103  
 INFORMATION FOR SEQ ID NO: 13:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 389 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-820-170A-13

Query Match 10.6%; Score 90; DB 2; Length 389;  
 Best Local Similarity 35.1%; Pred. No. 0.011;  
 Matches 27; Conservative 13; Mismatches 25; Indels 12; Gaps 3;

QY 16 KGLIYVPGTGSFLGAIANOLKS--KVRSTII--YLPEFIRTLKGFPGDSFEKK 70  
 DB 168 KGLIYVPGTGSFLGAIANOLKS--KVRSTII--YLPEFIRTLKGFPGDSFEKK 70  
 168 KGLIYVPGTGSFLGAIANOLKS--KVRSTII--YLPEFIRTLKGFPGDSFEKK 70

QY 71 LHRVREANITMLDDIGA 87  
 DB 221 YARDHOPCIIIFMEIDA 237

RESULT 10  
 US-09-055-699-13  
 Sequence 13, Application US/09055699  
 Patent No. 6005088

GENERAL INFORMATION:  
 APPLICANT: Teutomu, FUJIWARA  
 APPLICANT: Takeshi, WATANABE  
 APPLICANT: Masato, HORIE  
 APPLICANT: Toyomasa, KATAGIRI  
 TITLE OF INVENTION: HUMAN GENE  
 NUMBER OF SEQUENCES: 42  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas  
 STREET: 2100 Pennsylvania Avenue, N.W.  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: United States  
 ZIP: 20037-3202

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/055,699  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/820,170  
 FILING DATE:  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 293-7060  
 TELEFAX: (202) 293-7860  
 TELEX: 6491103  
 INFORMATION FOR SEQ ID NO: 13:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 389 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-09-055-699-13

Query Match 10.6%; Score 90; DB 3; Length 389;  
 Best Local Similarity 35.1%; Pred. No. 0.011;  
 Matches 27; Conservative 13; Mismatches 25; Indels 12; Gaps 3;

QY 16 KGLIYVPGTGSFLGAIANOLKS--KVRSTII--YLPEFIRTLKGFPGDSFEKK 70  
 DB 168 KGLIYVPGTGSFLGAIANOLKS--KVRSTII--YLPEFIRTLKGFPGDSFEKK 70  
 168 KGLIYVPGTGSFLGAIANOLKS--KVRSTII--YLPEFIRTLKGFPGDSFEKK 70

QY 71 LHRVREANITMLDDIGA 87  
 DB 221 YARDHOPCIIIFMEIDA 237

RESULT 11  
 US-09-273-565-13  
 Sequence 13, Application US/09273565A  
 Patent No. 6166190

GENERAL INFORMATION:  
 APPLICANT: FUJIWARA, TSUTOMU  
 APPLICANT: WATANABE, TAKESHI  
 APPLICANT: HORIE, MASATO  
 TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN  
 FILE REFERENCE: Q-53599  
 CURRENT APPLICATION NUMBER: US/09/273,565A  
 EARLIER FILING DATE: 1999-03-22  
 EARLIER APPLICATION NUMBER: 09/055,699  
 EARLIER FILING DATE: 1998-04-07  
 EARLIER APPLICATION NUMBER: 08/820,170  
 EARLIER FILING DATE: 1997-03-19  
 EARLIER APPLICATION NUMBER: JP 63410/1996  
 EARLIER FILING DATE: 1996-03-19  
 EARLIER APPLICATION NUMBER: JP 69163/1997  
 EARLIER FILING DATE: 1997-03-05  
 NUMBER OF SEQ ID NOS: 95  
 SOFTWARE: Patentin Ver. 2.1  
 SEQ ID NO 13  
 LENGTH: 389  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-273-565-13

Query Match 10.6%; Score 90; DB 3; Length 389;  
 Best Local Similarity 35.1%; Pred. No. 0.011;  
 Matches 27; Conservative 13; Mismatches 25; Indels 12; Gaps 3;

QY 16 KGLIYVPGTGSFLGAIANOLKS--KVRSTII--YLPEFIRTLKGFPGDSFEKK 70  
 DB 168 KGLIYVPGTGSFLGAIANOLKS--KVRSTII--YLPEFIRTLKGFPGDSFEKK 70  
 168 KGLIYVPGTGSFLGAIANOLKS--KVRSTII--YLPEFIRTLKGFPGDSFEKK 70

QY 71 LHRVREANITMLDDIGA 87  
 DB 221 YARDHOPCIIIFMEIDA 237

RESULT 12  
 US-09-565-538-13  
 Sequence 13, Application US/09565538  
 Patent No. 6333404  
 GENERAL INFORMATION:  
 APPLICANT: FUJIWARA, TSUTOMU  
 APPLICANT: WATANABE, TAKESHI  
 APPLICANT: HORIE, MASATO  
 TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN  
 FILE REFERENCE: Q-53599  
 CURRENT APPLICATION NUMBER: US/09/565,538  
 EARLIER FILING DATE: 2000-05-05  
 EARLIER APPLICATION NUMBER: 09/273,565  
 EARLIER FILING DATE: 1999-03-22  
 EARLIER APPLICATION NUMBER: 09/055,699  
 EARLIER FILING DATE: 1998-04-07  
 PRIOR APPLICATION NUMBER: 08/820,170

PRIOR FILING DATE: 1997-03-19  
PRIOR APPLICATION NUMBER: JP 63410/1996  
PRIOR FILING DATE: 1996-03-19  
PRIOR APPLICATION NUMBER: JP 69163/1997  
PRIOR FILING DATE: 1997-03-05  
NUMBER OF SEQ ID NOS: 95  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 13  
LENGTH: 389  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-565-538-13

Query Match 10.6%; Score 90; DB 4; Length 389;  
Best Local Similarity 35.1%; Pred. No. 0.011;  
Matches 27; Conservative 13; Mismatches 25; Indels 12; Gaps 3;

QY 16 KGLVYGPFGTGSFLLGAIANOLKS---KKVRSII--YLPEFIRTLKGFQDGSFEKK 70  
DB 168 KGLVYGPFGTGSFLLGAIANOLKS---KKVRSII--YLPEFIRTLKGFQDGSFEKK 70  
QY 71 LHRVREANILMDIGA 87  
DB 221 YARDHQPCTIFMDEIDA 237

RESULT 13  
US-09-661-468-13  
Sequence 13, Application US/09661468  
Patent No. 6376189  
GENERAL INFORMATION:  
APPLICANT: FUJIMURA, TSUTOMU  
APPLICANT: WATANABE, TAKESHI  
APPLICANT: HORIE, MASATO  
TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN  
FILE REFERENCE: Q-53599  
CURRENT APPLICATION NUMBER: US/09/661,468  
CURRENT FILING DATE: 2000-09-13  
PRIOR APPLICATION NUMBER: 09/055,699  
PRIOR FILING DATE: 1998-04-07  
PRIOR APPLICATION NUMBER: 08/820,170  
PRIOR FILING DATE: 1997-03-19  
PRIOR APPLICATION NUMBER: JP 63410/1996  
PRIOR FILING DATE: 1996-03-19  
PRIOR APPLICATION NUMBER: JP 69163/1997  
PRIOR FILING DATE: 1997-03-05  
NUMBER OF SEQ ID NOS: 95  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 13  
LENGTH: 389  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-661-468-13

Query Match 10.6%; Score 90; DB 4; Length 389;  
Best Local Similarity 35.1%; Pred. No. 0.011;  
Matches 27; Conservative 13; Mismatches 25; Indels 12; Gaps 3;

QY 16 KGLVYGPFGTGSFLLGAIANOLKS---KKVRSII--YLPEFIRTLKGFQDGSFEKK 70  
DB 168 KGLVYGPFGTGSFLLGAIANOLKS---KKVRSII--YLPEFIRTLKGFQDGSFEKK 70  
QY 71 LHRVREANILMDIGA 87  
DB 221 YARDHQPCTIFMDEIDA 237

RESULT 14  
US-09-976-165-13  
Sequence 13, Application US/09976165  
Patent No. 6562947  
GENERAL INFORMATION:

APPLICANT: FUJIMURA, TSUTOMU  
APPLICANT: WATANABE, TAKESHI  
APPLICANT: HORIE, MASATO  
TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN  
FILE REFERENCE: Q-53599  
CURRENT APPLICATION NUMBER: US/09/976,165  
CURRENT FILING DATE: 2001-10-15  
PRIOR APPLICATION NUMBER: 09/565,538  
PRIOR FILING DATE: 2000-05-05  
PRIOR APPLICATION NUMBER: 09/055,699  
PRIOR FILING DATE: 1998-04-07  
PRIOR APPLICATION NUMBER: 08/820,170  
PRIOR FILING DATE: 1997-03-19  
PRIOR APPLICATION NUMBER: JP 63410/1996  
PRIOR FILING DATE: 1996-03-19  
PRIOR APPLICATION NUMBER: JP 69163/1997  
PRIOR FILING DATE: 1997-03-05  
NUMBER OF SEQ ID NOS: 95  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 13  
LENGTH: 389  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-976-165-13

Query Match 10.6%; Score 90; DB 4; Length 389;  
Best Local Similarity 35.1%; Pred. No. 0.011;  
Matches 27; Conservative 13; Mismatches 25; Indels 12; Gaps 3;

QY 16 KGLVYGPFGTGSFLLGAIANOLKS---KKVRSII--YLPEFIRTLKGFQDGSFEKK 70  
DB 168 KGLVYGPFGTGSFLLGAIANOLKS---KKVRSII--YLPEFIRTLKGFQDGSFEKK 70  
QY 71 LHRVREANILMDIGA 87  
DB 221 YARDHQPCTIFMDEIDA 237

RESULT 15  
US-08-612-734B-2  
Sequence 2, Application US/08612734B  
Patent No. 5914246  
GENERAL INFORMATION:  
APPLICANT: Peery, Robert B.  
APPLICANT: Skatrud, Paul L.  
APPLICANT: Tobin, Matthew B.  
TITLE OF INVENTION: Multiple Drug Resistance Gene, of  
TITLE OF INVENTION: Aspergillus fumigatus  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Eli Lilly and Company  
STREET: Lilly Corporate Center, DC1501  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: U.S.A.  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/612,734B  
FILING DATE: 08-MAR-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Craig, Anne I.  
REGISTRATION NUMBER: 32,976  
REFERENCE/DOCKET NUMBER: X-9681  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-354-9570  
TELEFAX: 617-354-4043

; INFORMATION FOR SEQ ID NO: 2;  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1349 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-612-734B-2

Query Match 10.2%; Score 86.5; DB 2; Length 1349;  
Best Local Similarity 20.9%; Pred. No. 0.2;  
Matches 32; Conservative 27; Mismatches 49; Indels 45; Gaps 7;

Qy	3	DDICTAITNGEQVKGVLTYGPFGTGKSPILGAIANQLSKKVRSTIYLPPIRTLKGF	62
Db	456	EDVSLSMFAG---KTTALVGPSPSGSKSTVGLVER-----FYLPGGQVLLDG-	500
Qy	63	KDGSFEKKLHRVREANILMLDDIGAEEVTPWVRDEVIGPLHYRMVHELPTPPSNPDYS	122
Db	501	-----HDIQTLNL-----RWLRQOI-----SLVSOEPVLPSTTI-FR	531
Qy	123	ELSHHLAMTR-DGEKTKAARIIRVKSLSPTY	154
Db	532	NIEHGLIGTKFEHESKDKIRELVENAMARMANAH	564

Search completed: December 15, 2003, 15:17:39  
Job time : 8.57221 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 15:13:51 / Search time 16.0165 Seconds  
(without alignments)  
1904.368 Million cell updates/sec

Title: US-09-689-952-16  
Perfect score: 851  
Sequence: 1 AADICTAINTGEQVKGTYL.....ERVKSLSTPYFISGENFRNN 164

Scoring table: BLOSUM62  
Gapex 10.0, Gapext 0.5

Searched: 684280 seqs, 185983659 residues  
Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database:

Published Applications AA.\*  
1: /cgn2\_6/ptodata/1/pubppa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubppa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubppa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubppa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubppa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubppa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubppa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubppa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubppa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubppa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubppa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubppa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubppa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubppa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubppa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubppa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubppa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	104.5	12.3	248	12	US-10-287-274-435 Sequence 435, App
2	100	11.8	653	15	US-10-156-761-11850 Sequence 11850, A
3	98	11.5	467	9	US-09-815-242-10395 Sequence 10395, A
4	95.5	11.2	442	9	US-09-815-242-13552 Sequence 13552, A
5	92.5	10.8	574	10	US-09-764-868-774 Sequence 774, App
6	91.5	10.6	827	12	US-10-032-585-7621 Sequence 7621, App
7	90.5	10.6	454	9	US-09-815-242-11167 Sequence 11167, A
8	90	10.6	389	10	US-09-976-165-13 Sequence 13, App
9	90	10.6	389	12	US-10-205-219-165 Sequence 165, App
10	90	10.6	389	12	US-10-342-276-13 Sequence 13, App
11	90	10.6	393	15	US-10-128-714-3421 Sequence 3421, App
12	90	10.6	451	15	US-10-128-714-8421 Sequence 8421, App
13	88.5	10.4	796	12	US-10-032-585-7201 Sequence 7201, App
14	88	10.3	444	12	US-10-032-585-7385 Sequence 7385, App
15	88	10.3	453	9	US-09-815-242-5668 Sequence 5668, App

16	88	10.3	453	9	US-09-815-242-12701 Sequence 12701, A
17	87	10.2	411	12	US-10-032-585-7284 Sequence 7284, App
18	86.5	10.2	606	15	US-10-128-714-3064 Sequence 3064, App
19	86.5	10.2	847	15	US-10-128-714-8064 Sequence 8064, App
20	86	10.1	439	15	US-10-128-714-3522 Sequence 3522, App
21	86	10.1	439	15	US-10-128-714-8522 Sequence 8522, App
22	86	10.1	514	9	US-09-815-242-11772 Sequence 11772, A
23	86	10.1	1334	10	US-09-758-828-2 Sequence 2, App
24	85.5	10.0	444	15	US-10-153-668-308 Sequence 308, App
25	85.5	10.0	465	9	US-09-925-302-771 Sequence 771, App
26	85.5	10.0	482	10	US-09-764-868-1149 Sequence 1149, App
27	85.5	10.0	489	15	US-09-764-868-776 Sequence 776, App
28	85	10.0	421	15	US-10-128-714-8485 Sequence 8485, App
29	85	10.0	465	15	US-10-128-714-3485 Sequence 3485, App
30	84.5	9.9	442	11	US-09-828-062-7 Sequence 7, App
31	84	9.9	401	12	US-10-032-585-7325 Sequence 7325, App
32	84	9.9	561	15	US-10-156-761-13374 Sequence 13374, A
33	84	9.9	826	12	US-10-032-585-7227 Sequence 17, App
34	83.5	9.8	298	12	US-10-380-558-17 Sequence 21, App
35	83.5	9.8	456	9	US-09-756-998-21 Sequence 607, App
36	83	9.8	308	11	US-09-809-391-607 Sequence 607, App
37	83	9.8	308	12	US-09-882-171-607 Sequence 3369, App
38	83	9.8	389	15	US-10-128-714-3369 Sequence 3369, App
39	83	9.8	389	15	US-10-128-714-8369 Sequence 8369, App
40	83	9.8	399	11	US-09-809-391-474 Sequence 474, App
41	83	9.8	399	12	US-09-882-171-474 Sequence 474, App
42	83	9.8	728	10	US-09-712-363-167 Sequence 167, App
43	83	9.8	806	9	US-09-922-217-1117 Sequence 1117, App
44	83	9.8	806	9	US-09-922-217-1118 Sequence 1118, App
45	83	9.8	806	12	US-10-341-434-99 Sequence 99, App

#### ALIGNMENTS

RESULT 1  
US-10-287-274-435  
Sequence 435, Application US/10287274  
Publication No. US20030181408A1  
GENERAL INFORMATION:  
APPLICANT: Forsyth, R. Allyn  
APPLICANT: Ohlsen, Karl  
APPLICANT: Zyskind, Judith  
TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERETO  
FILE REFERENCE: ELIYRA, 008DVI  
CURRENT APPLICATION NUMBER: US/10/287,274  
CURRENT FILING DATE: 2002-10-31  
PRIOR APPLICATION NUMBER: US 60/164415  
PRIOR FILING DATE: 1999-11-09  
PRIOR APPLICATION NUMBER: US 09/711164  
NUMBER OF SEQ ID NOS: 469  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 435  
LENGTH: 248  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-10-287-274-435  
Query Match 12.3%; Score 104.5; DB 12; Length 248;  
Best Local Similarity 27.2%; Pred. No. 0.00078;  
Matches 40; Conservative 28; Mismatches 60; Indels 19; Gaps 6;  
QY 2 AADICTAINTGEQVKGTYLGPFGGKSTILGAINOLKSKVRSITITLPERIRLKGK 61  
DB 94 ADELMTGCTN-----FAFGSKPGTGKXNMLAIGRLKDKQTVVIVADVMXLRHS 147  
QY 62 FDGGSFEKKHRYREANTLMDDIGAEVTPWVDEVTGPIIHYVMVHELPPFSFND 120  
DB 148 YDGGSGGEFFLELCEVDLVADEIGIQRET---KNEQV-VLH-QIVRRIRASMSVGM 201  
QY 121 VSELEHILMTRDGEKTKVARIIRV 147

Db 202 LTNLNTYAMKTLIG-----RINDRM 222

RESULT 2

US-10-156-761-11850  
Sequence 11850, Application US/10156761  
Publication No. US20030119018A1

GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI

APPLICANT: IKEDA, HARUO

APPLICANT: ISHIKAWA, JUN

APPLICANT: HORIKAWA, HIROSHI

APPLICANT: SHIBA, TADAYOSHI

APPLICANT: SAKAKI, YOSHIYUKI

APPLICANT: HATTORI, MASAHIRA

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-262

CURRENT APPLICATION NUMBER: US/10/156,761

PRIOR FILING DATE: 2002-05-29

PRIOR APPLICATION NUMBER: JP 2001-204089

PRIOR FILING DATE: 2001-05-30

PRIOR APPLICATION NUMBER: JP 2001-272697

PRIOR FILING DATE: 2001-08-02

NUMBER OF SEQ ID NOS: 15109

SEQ ID NO 11850

LENGTH: 653

TYPE: PR

ORGANISM: Streptomyces avermitilis

US-10-156-761-11850

Query Match 11.8%; Score 100; DB 15; Length 653;  
Best Local Similarity 39.2%; Pred. No. 0.01;

Matches 29; Conservative 15; Mismatches 20; Indels 10; Gaps 4;

QY 18 LTVYGPFGKSFILGAINOLKSKKVSSTIYLP--PEPRTLKGGFKDGSFEKKLHVR 71

DB 350 LFTYGGPGIGKTHLNAIGNAILKNI PNARVXIPAESFINDPLDHLRLGEMEKFKTYR 406

QY 72 HVRVRENIIMLDI 85

DB 407 -RRYEMDILLVDI 419

RESULT 3

US-09-815-242-10395

Sequence 10395, Application US/09815242

Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari L.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

TITLE OF INVENTION: Prokaryotes

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

PRIOR FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR FILING DATE: 2000-12-22

PRIOR FILING DATE: 2000-12-22

PRIOR FILING DATE: 2000-12-22

PRIOR FILING DATE: 2000-12-22

PRIOR FILING DATE: 2000-12-22

PRIOR FILING DATE: 2000-12-22

PRIOR FILING DATE: 2000-12-22

PRIOR FILING DATE: 2000-12-22

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 10395

LENGTH: 467

TYPE: PR

ORGANISM: Escherichia coli

US-09-815-242-10395

Query Match 11.5%; Score 98; DB 9; Length 467;  
Best Local Similarity 27.1%; Pred. No. 0.011;

Matches 19; Conservative 21; Mismatches 28; Indels 2; Gaps 1;

QY 18 LTVYGPFGKSFILGAINOLKSKKVSSTIYLP--PEPRTLKGGFKDGSFEKKLHVR 75

DB 168 LFTYGGPGIGKTHLNAIGNAILKNI PNARVXIPAESFINDPLDHLRLGEMEKFKTYR 227

QY 76 EANIIMLDI 85

DB 228 SVDALIIDDI 237

RESULT 4

US-09-815-242-13552

Sequence 13552, Application US/09815242

Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari L.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

TITLE OF INVENTION: Prokaryotes

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

PRIOR FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 13552

LENGTH: 442

TYPE: PR

ORGANISM: Streptococcus pneumoniae

US-09-815-242-13552

Query Match 11.2%; Score 95.5; DB 9; Length 442;  
Best Local Similarity 23.8%; Pred. No. 0.02;

Matches 29; Conservative 22; Mismatches 54; Indels 17; Gaps 2;

QY 18 LTVYGPFGKSFILGAINOLKSKKVSSTIYLP--PEPRTLKGGFKDGSFEKKLHVR 75

DB 140 LFTYGGPGIGKTHLNAIGNAILKNI PNARVXIPAESFINDPLDHLRLGEMEKFKTYR 199

QY 76 EANIIMLDIGAEBVTWVDEVIQPLLHYRWVHELPTFFSSNPDYSLEHHLAATRDGE 135

DB 76 EANIIMLDIGAEBVTWVDEVIQPLLHYRWVHELPTFFSSNPDYSLEHHLAATRDGE 135

QY 76 EANIIMLDIGAEBVTWVDEVIQPLLHYRWVHELPTFFSSNPDYSLEHHLAATRDGE 135

DB 76 EANIIMLDIGAEBVTWVDEVIQPLLHYRWVHELPTFFSSNPDYSLEHHLAATRDGE 135

QY 76 EANIIMLDIGAEBVTWVDEVIQPLLHYRWVHELPTFFSSNPDYSLEHHLAATRDGE 135

DB 76 EANIIMLDIGAEBVTWVDEVIQPLLHYRWVHELPTFFSSNPDYSLEHHLAATRDGE 135

Db 200 SJDLILDDIOS-----LSGKVVATQEEFFNTFNALHDKQKQIVLTSDRS 244

Qy 136 BK 137  
Db 245 PK 246

RESULT 5  
US-09-764-868-774

; Sequence 774, Application US/09764868  
; Patent No. US20020168711A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PT232

; CURRENT APPLICATION NUMBER: US/09/764,868

; PRIOR FILING DATE: 2001-01-17

; PRIOR APPLICATION DATA REMOVED - refer to PALM or file wrapper

; NUMBER OF SEQ ID NOS: 1510

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 774

; LENGTH: 574

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-764-868-774

Query Match 10.9%; Score 92.5; DB 10; Length 574;  
Best Local Similarity 30.3%; Pred. No. 0.064;  
Matches 30; Conservative 17; Mismatches 33; Indels 19; Gaps 4;

Qy 16 KSLVLYGPGTGSFTLGAIAIOLKSKVRSITLYLPEFRTLKGFGKDSFEKILHRV-74

Db 127 RGVLYGPGTGTGKTMLARAVANEVGA---YVSINGPETISIKRY-----GETRAKLRQIF 178

Qy 75 -----REANILMLDDIGA-----EAVTPWVRDEVIGPL 103

Db 179 AEATLHPSIIFIDELDALCPKREGAQNEVEKRVASLL 217

RESULT 6  
US-10-032-585-7621

; Sequence 7621, Application US/10032585

; Publication No. US20030180953A1

; GENERAL INFORMATION:

; APPLICANT: Terry, Roemer D.

; APPLICANT: Bo, Jiang

; APPLICANT: Charles, Boone

; APPLICANT: Howard, Bussey

; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery

; FILE REFERENCE: 10182-005-999

; CURRENT APPLICATION NUMBER: US/10/032,585

; PRIOR FILING DATE: 2001-12-20

; NUMBER OF SEQ ID NOS: 8000

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 7621

; LENGTH: 827

; TYPE: PRT

; ORGANISM: Candida albicans

US-10-032-585-7621

Query Match 10.8%; Score 91.5; DB 12; Length 827;  
Best Local Similarity 31.8%; Pred. No. 0.14; Indels 17; Gaps 6;  
Matches 35; Conservative 17; Mismatches 41; Indels 17; Gaps 6;

Qy 10 TNGEQVKGVLVYGPFGTGSFTLGAIAIOLKSKVRSITLYLPEFRTLKGFGKDSFEK 69

Db 222 TGEVPRGVLYGPGTGTGKTMLARAVANEVGA---ISASVSGM-----SGESSEK 273

Qy 70 KLRV-REAN-----ILMLDDIGAEEVTPWVRDEVIGPLHVRMHELPT 113

Db 274 KLRVREANQIAPCLIFMDEIDA--ITP-KRDGGAQREMERKRIVALLLT 320

RESULT 7  
US-09-815-242-1167

; Sequence 1167, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Karl L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1167

; LENGTH: 454

; TYPE: PRT

; ORGANISM: Haemophilus influenzae

US-09-815-242-1167

Query Match 10.6%; Score 90.5; DB 9; Length 454;  
Best Local Similarity 26.8%; Pred. No. 0.08;  
Matches 37; Conservative 25; Mismatches 47; Indels 29; Gaps 7;

Qy 19 YLYGPGTGSFTLGAIAIOLKSKVRSITLYLPEFRTLKGFGKDSFEK-KLHRV 74

Db 157 FLYGPGTGSFTLGAIAIOLKSKVRSITLYLPEFRTLKGFGKDSFEK-KLHRV 216

Qy 75 REANILMLDDI-----GAEEVTPWVRDEVIGPLHVRMHELPTFSSNPFYSELEHILANT 131

Db 217 LDA--LLVDDIQFPAKEKTKQ-----BEFFHIFNSLF--ETGRQIILT 255

Qy 132 RDGEERTKARILIERVKS 149

Db 256 SDRVPK-EIEKIEERLKS 272

RESULT 8  
US-09-976-165-13

; Sequence 13, Application US/09976165

; Patent No. US20020107383A1

; GENERAL INFORMATION:

; APPLICANT: FUJIMURA, TSUTOMU

; APPLICANT: WATANABE, TAKESHI

; APPLICANT: HORIE, MASATO

; TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN

; FILE REFERENCE: O-53599

; CURRENT APPLICATION NUMBER: US/09/976,165

; PRIOR FILING DATE: 2001-10-15

; PRIOR APPLICATION NUMBER: 09/565,538

; PRIOR FILING DATE: 2000-05-05

; PRIOR APPLICATION NUMBER: 09/055,699

;; PRIOR FILING DATE: 1998-04-07  
;; PRIOR APPLICATION NUMBER: 08/820,170  
;; PRIOR FILING DATE: 1997-03-19  
;; PRIOR APPLICATION NUMBER: JP 63410/1996  
;; PRIOR FILING DATE: 1996-03-19  
;; PRIOR APPLICATION NUMBER: JP 69163/1997  
;; PRIOR FILING DATE: 1997-03-05  
;; NUMBER OF SEQ ID NOS: 95  
;; SOFTWARE: Patentin Ver. 2.1  
;; SEQ ID NO: 13  
;; LENGTH: 389  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-976-165-13

Query Match 10.6%; Score 90; DB 10; Length 389;  
Best Local Similarity 35.1%; Pred. No. 0.073;  
Matches 27; Conservative 13; Mismatches 25; Indels 12; Gaps 3;

QY 16 KGLYLGPFGTGSFLIGAIANOLKS---KKVRSTII--YLPEFIRTLKGGFQDGSFEKK 70  
DB 168 KGCILVGPFGTGTLLARAVASQDCNFKLVSSIVDKYIGSARLIREMF-----N 220  
71 LHRVREANILMLDDIGA 87  
DB 221 YARDHQPCTIFMDEIDA 237

RESULT 9  
US-10-205-219-165  
;; Sequence 165, Application US/10205219  
;; Publication No. US20030138803A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Warner-Lambert Company  
;; APPLICANT: Lee, Kevin  
;; APPLICANT: Dixon, Alister  
;; APPLICANT: Brookbank, Robert  
;; APPLICANT: Plimock, Robert  
;; TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain  
;; FILE REFERENCE: WU-A-018200  
;; CURRENT APPLICATION NUMBER: US/10/205,219  
;; CURRENT FILING DATE: 2002-07-24  
;; PRIOR APPLICATION NUMBER: GB 0118354.0  
;; PRIOR FILING DATE: 2001-07-27  
;; NUMBER OF SEQ ID NOS: 197  
;; SOFTWARE: Patentin Ver. 2.1  
;; SEQ ID NO: 165  
;; LENGTH: 389  
;; TYPE: PRT  
;; ORGANISM: Spentrophilus tridecemlineatus  
;; FEATURE:  
;; OTHER INFORMATION: ATPase domain protein 44  
US-10-205-219-165

Query Match 10.6%; Score 90; DB 12; Length 389;  
Best Local Similarity 35.1%; Pred. No. 0.073;  
Matches 27; Conservative 13; Mismatches 25; Indels 12; Gaps 3;

QY 16 KGLYLGPFGTGSFLIGAIANOLKS---KKVRSTII--YLPEFIRTLKGGFQDGSFEKK 70  
DB 168 KGCILVGPFGTGTLLARAVASQDCNFKLVSSIVDKYIGSARLIREMF-----N 220  
71 LHRVREANILMLDDIGA 87  
DB 221 YARDHQPCTIFMDEIDA 237

RESULT 10  
US-10-342-276-13  
;; Sequence 13, Application US/10342276  
;; Publication No. US20030143688A1  
;; GENERAL INFORMATION:  
;; APPLICANT: FUJIMURA, TSUTOMU

;; APPLICANT: WATANABE, TAKESHI  
;; APPLICANT: HORIE, MASATO  
;; TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN  
;; TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIOUITIN-CONUGATING ENZYME  
;; FILE REFERENCE: Q-53599  
;; CURRENT APPLICATION NUMBER: US/10/342,276  
;; CURRENT FILING DATE: 2003-01-15  
;; PRIOR APPLICATION NUMBER: US/09/976,165  
;; PRIOR FILING DATE: 2001-10-15  
;; PRIOR APPLICATION NUMBER: 09/565,538  
;; PRIOR FILING DATE: 2000-05-05  
;; PRIOR APPLICATION NUMBER: 09/055,699  
;; PRIOR FILING DATE: 1998-04-07  
;; PRIOR APPLICATION NUMBER: 08/820,170  
;; PRIOR FILING DATE: 1997-03-19  
;; PRIOR APPLICATION NUMBER: JP 63410/1996  
;; PRIOR FILING DATE: 1996-03-19  
;; PRIOR APPLICATION NUMBER: JP 69163/1997  
;; PRIOR FILING DATE: 1997-03-05  
;; NUMBER OF SEQ ID NOS: 95  
;; SOFTWARE: Patentin Ver. 2.1  
;; SEQ ID NO: 13  
;; LENGTH: 389  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-342-276-13

Query Match 10.6%; Score 90; DB 12; Length 389;  
Best Local Similarity 35.1%; Pred. No. 0.073;  
Matches 27; Conservative 13; Mismatches 25; Indels 12; Gaps 3;

QY 16 KGLYLGPFGTGSFLIGAIANOLKS---KKVRSTII--YLPEFIRTLKGGFQDGSFEKK 70  
DB 168 KGCILVGPFGTGTLLARAVASQDCNFKLVSSIVDKYIGSARLIREMF-----N 220  
71 LHRVREANILMLDDIGA 87  
DB 221 YARDHQPCTIFMDEIDA 237

RESULT 11  
US-10-128-714-3421  
;; Sequence 3421, Application US/10128714  
;; Publication No. US20030119013A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Jiang, Bo  
;; APPLICANT: Hu, Weng  
;; APPLICANT: Tishkoff, Daniel  
;; APPLICANT: Zamudio, Carlos  
;; APPLICANT: Broshkin, Alexey M  
;; APPLICANT: Lemieux, Sebastien M  
;; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and  
;; TITLE OF INVENTION: Methods of Use  
;; FILE REFERENCE: 10182-018-999  
;; CURRENT APPLICATION NUMBER: US/10/128,714  
;; CURRENT FILING DATE: 2002-04-23  
;; PRIOR APPLICATION NUMBER: US 60/285,697  
;; PRIOR FILING DATE: 2001-04-23  
;; PRIOR APPLICATION NUMBER: US 60/287,066  
;; PRIOR FILING DATE: 2001-04-27  
;; PRIOR APPLICATION NUMBER: US 60/295,890  
;; PRIOR FILING DATE: 2001-06-05  
;; PRIOR APPLICATION NUMBER: US 60/303,899  
;; PRIOR FILING DATE: 2001-07-09  
;; PRIOR APPLICATION NUMBER: US 60/316,362  
;; PRIOR FILING DATE: 2001-08-31  
;; NUMBER OF SEQ ID NOS: 8603  
;; SOFTWARE: Patentin version 3.1  
;; SEQ ID NO: 3421  
;; LENGTH: 393  
;; TYPE: PRT  
;; ORGANISM: Aspergillus fumigatus  
US-10-128-714-3421

APPLICANT:	Yamamoto, Robert T
APPLICANT:	Carr, Grant J.
APPLICANT:	YU, H. Howard

```
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5668
; LENGTH: 453
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-09-815-242-5668

Query Match      10.3%; Score 88; DB 9; Length 453;
Best Local Similarity 28.4%; Pred. No. 0.16;
Matches 21; Conservative 17; Mismatches 26; Indels 10; Gaps 2;

QY      18 LVYGPFGTGSFLGAIANQLSKKVRSTIYLP-----EPIRTLKGGFKDGSPEKKL 71
      152 LFTYGGVGLGKTHLMHAIGHVLDNNPDAKVITYTSEKFTNEFIKSI---RDNKGGAFR 207
      208 ERYRNIDVLLIDDI 221

QY      72 HRYRANIIMLDDI 85
      208 ERYRNIDVLLIDDI 221

Db
```

Search completed: December 15, 2003, 15:19:02  
Job time : 16.0165 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 15:11:40 ; Search time 9.24897 Seconds  
(without alignments)  
1705.235 Million cell updates/sec

Title: US-09-689-952-16

Perfect score: 851

Sequence: 1 AADDICTAITNGEQVKGGLYL.....BRVKSLSPTPYFLSGENPRNN 164

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: PIR76:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	851	100.0	306	2 A89952	primosomal protein
2	427.5	50.2	311	1 IOBS44	primosome componen
3	423.5	49.8	307	2 AB1632	primosome componen
4	420.5	49.4	311	2 H84042	primosome componen
5	418.5	49.2	307	2 AH1269	primosome componen
6	347	40.8	298	2 B98066	primosome componen
7	346	40.7	298	2 D95199	primosomal protein
8	330	38.8	293	2 H86718	primosomal protein
9	166	19.5	209	2 S77882	dnab protein homol
10	140.5	16.5	282	2 D97138	DNA replication pr
11	140	16.5	316	2 B82807	conserved hypotet
12	129.5	15.2	235	2 E70378	DNA replication pr
13	127.5	15.0	313	2 D69945	phage-related prot
14	127	14.9	266	2 I40411	PBSX prophage ORF
15	125.5	14.7	440	2 D81415	chromosomal repli
16	114	13.4	265	1 BVEICIT	1stb protein - Esc
17	111	13.0	350	2 B97735	probable atpase n2
18	109	12.8	316	2 T10436	probable transpos
19	105.5	12.4	248	2 C85626	probable DNA repli
20	105.5	12.4	248	2 B90763	probable DNA repli
21	105	12.3	246	2 C90913	probable DNA repli
22	104.5	12.3	248	1 C64886	DNA replication pr
23	104.5	12.3	248	2 A85477	probable DNA repli
24	104.5	12.3	248	2 A99872	probable DNA repli
25	104.5	12.3	287	2 AD1888	similarities Scap
26	104.5	12.3	287	2 AG1651	similarities Staph
27	104	12.2	520	2 H69125	hypothetical prote
28	103.5	12.2	248	2 AC0738	probable DNA repli
29	103.5	12.2	263	2 T47074	hypothetical prote

30	103.5	12.2	726	2 D72613	probable transito
31	103	12.1	454	2 D84931	chromosomal replic
32	102.5	12.0	246	2 D84932	DNA replication pr
33	102.5	12.0	432	2 S62461	probable AAA-famil
34	102.5	12.0	783	2 T18421	hypothetical prote
35	101.5	11.9	733	2 H69411	cell division cont
36	101	11.9	251	2 S23889	hypothetical prote
37	101	11.9	294	2 C66802	hypothetical prote
38	101	11.9	699	2 B72479	probable transito
39	100.5	11.8	357	2 D71732	probable atpase n2
40	100.5	11.8	811	2 B69512	cell division cont
41	100	11.8	291	2 A86681	hypothetical prote
42	100	11.8	307	2 G64239	Holliday junction
43	100	11.8	462	2 AG0497	chromosomal replic
44	100	11.8	656	2 A41870	dnab protein - Str
45	99	11.6	466	1 IOBSV	replication initia

#### ALIGNMENTS

RESULT 1  
A89952  
primosomal protein [imported] - Staphylococcus aureus (strain N315)  
C/Species: Staphylococcus aureus  
C/Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C/Accession: A89952  
R/Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; Ito, C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A/Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A/Reference number: A89758; MUID:21311952; PMID:11418146  
A/Accession: A89952  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-306 <KUR>  
A/Cross-references: GB:BA000018; PID:g13701480; PIDN:BA842774.1; GSPDB:GN00149  
A/Experimental source: strain N315  
C/Genetics:  
A/Genes: dnaI  
C/Superfamily: 44K dnab protein homolog

Query Match 100.0%; Score 851; DB 2; Length 306;  
Best Local Similarity 100.0%; Pred. No. 5.4e-68;  
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	AADDICTAITNGEQVKGGLYGPFGTCKSFILGAIANOLSKKVRSTIIYLPFIRTLKG	60
DB	143	AADDICTAITNGEQVKGGLYVPGFTGKSFILGAIANOLSKKVRSTIIYLPFIRTLKG	202
QY	61	GFKDSFEKKLRHREANILIMDDIGAEVTPWVDEVIQPLIHYRMVHELPTFPSSND	120
DB	203	GFKDSFEKKLRHREANILIMDDIGAEVTPWVDEVIQPLIHYRMVHELPTFPSSND	262
QY	121	YSELHHLAMTRDGEKTKAARIIRVKSLSPTPYFLSGENPRNN	164
DB	263	YSELHHLAMTRDGEKTKAARIIRVKSLSPTPYFLSGENPRNN	306

RESULT 2  
IOBS44  
primosome component (helicase loader) dnaI - Bacillus subtilis  
N/A/Alternate names: dnab protein homolog, 44K, hypothetical protein Y (dnab 3' region)  
C/Species: Bacillus subtilis  
C/Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 19-Jan-2001  
C/Accession: B24720; C26580; F69617  
R/Ogasawara, N.; Moriya, S.; Mazza, P. G.; Yoshikawa, H.  
Nucleic Acids Res. 14, 9989-9999, 1986  
A/Title: Nucleotide sequence and organization of dnab gene and neighbouring genes on the  
A/Reference number: A93650; MUID:87117549; PMID:3027671  
A/Accession: B24720  
A/Molecule type: DNA

A/Residues: 1-311 <OGA>  
 A/Cross-references: GB:X04963; NID:G39880; PIDN:CAA2863.1; PID:G39881  
 R/Hoshino, T.; McKenzie, T.; Schmidt, S.; Tanaka, T.; Sueoka, N.  
 Proc. Natl. Acad. Sci. U.S.A. 84, 653-657, 1987  
 A/Title: Nucleotide sequence of *Bacillus subtilis* dnaB: a gene essential for DNA replication  
 A/Reference number: A94709; MUID:87118226; PMID:3027697  
 A/Accession: C26580  
 A/Molecule type: DNA  
 A/Residues: 1-18, 'N', 20-23, 'T', 25-206 <HOS>  
 R/Kunet, F.; Ogaawara, N.; Moszer, I.; Albertini, A.M.; Allion, G.; Azevedo, V.; Berche, C.; Bron, S.; Brouillet, S.; Busch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E. Nature 390, 249-256, 1997  
 A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Galliz, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holzapfel, S.; Hosono, S.; Hullo, M.F.; Koetter, P.; Koningslee, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lathionis, A.; Authors: Lander, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetello, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, A.; Authors: Schleicher, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot, Akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terstra, P.; Tognoni, A.; Toasato, V.; Uchiyama, T.; Winters, P.; Wipet, A.; Yamamoto, H.; Yamane, K.; Yaeumoto, K.; Yata, K.; Yoshida, K.  
 A/Authors: Yoshikawa, H.F.; Zundstein, E.; Yoshikawa, H.; Danchin, A.  
 A/Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
 A/Reference number: A69580; MUID:98044033; PMID:9384377  
 A/Accession: F69617  
 A/Status: nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-311 <KUN>  
 A/Cross-references: GB:Z99118; GB:AL009126; NID:G2635200; PIDN:CAB14858.1; PID:G2635363  
 A/Experimental source: strain 168  
 C/Genetics:  
 A/Gene: dnaI  
 C/Superfamily: 44k dnaa protein homolog  
 C/Keywords: ATP; nucleotide binding; P-loop  
 F:168-175/Region: nucleotide-binding motif A (P-loop)  
 Query Match 50.2%; Score 427.5; DB 1; Length 311;  
 Best Local Similarity 55.7%; Pred. No. 1.9e-30;  
 Matches 83; Conservative 26; Mismatches 39; Indels 1; Gaps 1;  
 QY 16 KGLYVGFPGTGSFLLGAIANOLSKKVRSTIYLPDEFRTLKGFQDGFSEKGLHRRV 75  
 DB 162 KGLYVGFPGTGSFLLGAIANOLSKKVRSTIYLPDEFRTLKGFQDGFSEKGLHRRV 221  
 QY 76 EANIMLDDIGAEVTPWVRDEVIGPLHYRWVHELPFFSSNPDYSELEHILAMTDE 135  
 DB 222 TTPVLMDDIGAEVTPWVRDEVIGPLHYRWVHELPFFSSNPDYSELEHILAMTDE 281  
 QY 136 -EKTAKARIIRVKSSTPYFLSGENFRN 163  
 DB 282 KEVKAKARIIRVKSSTPYFLSGENFRN 310  
 RESULT 3  
 ABI632  
 primosome component (helicase loader) DnaI [imported] - *Listeria innocua* (strain Clp112  
 C/Species: *Listeria innocua*  
 C/Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
 C/Accession: ABI632  
 R/Glaeser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Feihl, H.; Jones, L.M.; Karst, U. Science 294, 849-852, 2001  
 A/Authors: Kreft, U.; Kuhn, M.; Kunst, F.; Kurapk, G.; Madueno, E.; Matounam, A.; Maok, C.; Schluter, T.; Simoes, N.; Tietz, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend, A./Title: Comparative genomics of *Listeria* species.  
 A/Reference number: AB1077; MUID:21537279; PMID:11679669  
 A/Accession: ABI632  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-307 <GUA>  
 A/Cross-references: GB:AL592022; PIDN:CAC96826.1; PID:G16414082; GSPDB:GN00178

A/Experimental source: strain Clp11262  
 C/Genetics:  
 A/Gene: dnaI  
 C/Superfamily: 44k dnaa protein homolog  
 Query Match 49.8%; Score 423.5; DB 2; Length 307;  
 Best Local Similarity 52.3%; Pred. No. 4.3e-30;  
 Matches 80; Conservative 32; Mismatches 40; Indels 1; Gaps 1;  
 QY 11 NGRVYVGLYVGFPGTGSFLLGAIANOLSKKVRSTIYLPDEFRTLKGFQDGFSEK 70  
 DB 154 NGRVYVGLYVGFPGTGSFLLGAIANOLSKKVRSTIYLPDEFRTLKGFQDGFSEK 213  
 QY 71 LHRVREANIMLDDIGAEVTPWVRDEVIGPLHYRWVHELPFFSSNPDYSELEHILAM 130  
 DB 214 IQPARKTEVLMDDIGAEVTPWVRDEVIGPLHYRWVHELPFFSSNPDYSELEHILAM 273  
 QY 131 TRDG-EKTAKARIIRVKSSTPYFLSGENFRN 162  
 DB 274 AQNGTEKAKARIIRVKSSTPYFLSGENFRN 306  
 RESULT 4  
 H84042  
 primosome component (helicase loader) DnaI [imported] - *Bacillus halodurans* (strain C-12;  
 C/Species: *Bacillus halodurans*  
 C/Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
 C/Accession: H84042  
 R/Takaki, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Maui, N.; Fuji, F.; Hiran  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A/Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and  
 A/Reference number: A83650; MUID:20512582; PMID:11058132  
 A/Accession: H84042  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-311 <STO>  
 A/Cross-references: GB:AP001517; GB:BA000004; NID:G10175500; PIDN:BA06863.1; GSPDB:GN001  
 A/Experimental source: strain C-125  
 C/Genetics:  
 A/Gene: dnaI  
 C/Superfamily: 44k dnaa protein homolog  
 Query Match 49.4%; Score 420.5; DB 2; Length 311;  
 Best Local Similarity 50.0%; Pred. No. 8e-30;  
 Matches 77; Conservative 33; Mismatches 43; Indels 1; Gaps 1;  
 QY 12 GEGVYVGLYVGFPGTGSFLLGAIANOLSKKVRSTIYLPDEFRTLKGFQDGFSEK 71  
 DB 156 GEGVYVGLYVGFPGTGSFLLGAIANOLSKKVRSTIYLPDEFRTLKGFQDGFSEK 215  
 QY 72 HRYREANIMLDDIGAEVTPWVRDEVIGPLHYRWVHELPFFSSNPDYSELEHILAMT 131  
 DB 216 DFKVNAQVLPFDIGAEVTPWVRDEVIGPLHYRWVHELPFFSSNPDYSELEHILAMT 275  
 QY 132 -RDGEKTAARIIRVKSSTPYFLSGENFRN 164  
 DB 276 DKSGTEKAKARIIRVKSSTPYFLSGENFRN 309  
 RESULT 5  
 AH1269  
 primosome component (helicase loader) DnaI [imported] - *Listeria monocytogenes* (strain EC  
 C/Species: *Listeria monocytogenes*  
 C/Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
 C/Accession: AH1269  
 R/Glaeser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Feihl, H.; Jones, L.M.; Karst, U. Science 294, 849-852, 2001  
 A/Authors: Kreft, U.; Kuhn, M.; Kunst, F.; Kurapk, G.; Madueno, E.; Matounam, A.; Maok, C.; Schluter, T.; Simoes, N.; Tietz, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend, A./Title: Comparative genomics of *Listeria* species.  
 A/Reference number: AB1077; MUID:21537279; PMID:11679669

dnaA protein homolog - *Mycoplasma capricolum*

N/Alternate names: protein MC072  
 C/Species: Mycoplasma capricolum  
 C/Date: 09-Oct-1997 #sequence\_revision 24-Oct-1997 #text\_change 07-Dec-1999  
 C/Accession: S77882  
 R/Bork, P.; Ouzounis, C.; Casari, G.; Schneider, R.; Sander, C.; Dolan, M.; Gilbert, W.;  
 Mol. Microbiol. 16, 955-967, 1995  
 A/Title: Exploring the Mycoplasma capricolum genome: a minimal cell reveals its phylotoic  
 A/Reference number: S77739; MUID:9605641; PMID:7476192  
 A/Accession: S77882  
 A/Status: nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-209 <BCR>  
 A/Cross-references: EMBL:Z33058; NID:G514450; PIDN:CAA83732.1; PID:G530419  
 A/Experimental source: ATCC 27343  
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994  
 C/Genetics:  
 A/Genetic code: SGC3

Query Match 19.5%; Score 166; DB 2; Length 209;  
 Best Local Similarity 35.9%; Pred. No. 1.8e-07;  
 Matches 37; Conservative 21; Mismatches 39; Indels 6; Gaps 3;  
 QY 16 KGLVLPFGTSGFLLGAIANOLSKKVRSTIY--LPEFIRTLKGGFGK--KTL 71  
 DB 100 KGVLYGHSGIGKTYMKVLANLASK--NKTVIFSTRSLIDRLKESFNSSEINSLM 157  
 QY 72 HRREANIIMLDDIGAEVTPWVRDEVIGPLHYRMVHELPTF 114  
 DB 158 KKIKTVDPLFLDDIGENLSLWADPLFEVLANRMENOKATP 200

## RESULT 10

D97138  
 DNA replication protein Dnac [imported] - Clostridium acetobutylicum  
 C/Species: Clostridium acetobutylicum  
 C/Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
 C/Accession: D97138  
 R/Daly, M.J.; Bennett, G.N.; Koehn, E.V.; Smith, D.R.  
 J. Bacteriol. 183, 4823-4838, 2001  
 A/Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld  
 A/Reference number: A96900; MUID:21359325; PMID:21359325  
 A/Accession: D97138  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-282 <KRR>  
 A/Cross-references: GB:AE001437; PIDN:AAK79895.1; PID:G15024913; GSPDB:GN00168  
 A/Experimental source: Clostridium acetobutylicum ATCC824  
 C/Genetics:  
 A/Genetic code: CAC1933

Query Match 16.5%; Score 140.5; DB 2; Length 282;  
 Best Local Similarity 23.5%; Pred. No. 4.7e-05;  
 Matches 36; Conservative 40; Mismatches 68; Indels 9; Gaps 3;  
 QY 13 EOVKGLVLPFGTSGFLLGAIANOLSKKVRSTIY--LPEFIRTLKGGFGKDSFEKKH 72  
 DB 120 QQGLGLIYSGAGNKYTLASALANELLKQYIPVVCVINGLSRIQKTYNSNGKEASD 179  
 QY 73 RVR---EANTIMLDDIGAEVTPWVRDEVIGPLHYRMVHELPTFFSSNFDYSELEHHA 129  
 DB 180 IISFIVADLLIIDLCTEKSEMSR-SMTYINVDSSYRSKLPILITNSLEINPSKRGHV 238  
 QY 130 MTRDGEKTKAARIIEVKSLSTPYLGENFR 162  
 DB 239 IADQYHERTES-----RIFEMCTPVENTSKSIR 266

## RESULT 11

B82907  
 conserved hypothetical ATP/GTP-binding protein UU307 [imported] - Ureaplasma urealyticum  
 C/Species: Ureaplasma urealyticum  
 C/Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000

C/Accession: B82907  
 R/Glass, J.I.; Ieflowicz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Casseil, G.H.  
 Submitted to Genbank, February 2000  
 A/Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a mini  
 A/Reference number: A82870  
 A/Accession: B82907  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-316 <GLA>  
 A/Cross-references: GB:AE002128; GB:AF222894; NID:G6899279; PIDN:AAF30716.1; GSPDB:GN001;  
 A/Experimental source: setovar 3; biovar 1  
 C/Genetics:  
 A/Genetic code: SGC3

Query Match 16.5%; Score 140; DB 2; Length 316;  
 Best Local Similarity 25.5%; Pred. No. 6e-05;  
 Matches 39; Conservative 28; Mismatches 68; Indels 18; Gaps 3;  
 QY 11 NGBQVGLVLPFGTSGFLLGAIANOLSKKVRSTIY--LPEFIRTLKGGFK----- 63  
 DB 151 NNQPSAFIYGFDSVGSIIITQATNTI--SLKTNLKIAYT-----TLNDFKRVYQFEN 204  
 QY 64 -----DSEFEKTLHVRREANIIMLDDIGAEVTPWVRDEVIGPLHYRMVHELPTFFSSN 118  
 DB 205 YKOTSDVLVINELVNLSNDVLDVIDPSSVNLNWSISITLWPIENRLKSTRTQITFISN 264  
 QY 119 FDYSLEHHLAMTRDGEKTKAARIIEVKSL 151  
 DB 265 FSIEQLINSTKNTTNEQTKLRLFRRIEYL 297

## RESULT 12

E70378  
 DNA replication protein Dnac - Aquifex aeolicus  
 C/Species: Aquifex aeolicus  
 C/Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 05-Nov-1999  
 C/Accession: E70378  
 R/Dackert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove  
 V.  
 Nature 392, 353-358, 1998  
 A/Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
 A/Reference number: A70300; MUID:98196666; PMID:9537320  
 A/Accession: E70378  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-235 <AOF>  
 A/Cross-references: GB:AE000713; NID:G2983424; PIDN:AACT07013.1; PID:G2983431; GB:AE00065;  
 A/Experimental source: strain VFS  
 C/Genetics:  
 A/Genetic code: dnac

Query Match 15.2%; Score 129.5; DB 2; Length 235;  
 Best Local Similarity 26.5%; Pred. No. 0.0006;  
 Matches 45; Conservative 27; Mismatches 61; Indels 37; Gaps 5;  
 QY 11 NGBQVGLVLPFGTSGFLLGAIANOLSKK--VRSTIY--LPEFIRTLKGGFGKDSFEK 69  
 DB 75 NPEBGKGLTFVSGPGVKTHLAVATLKAIEKKGRNGYFPDTRDLFRKHLMDCKDK 134  
 QY 70 KLRHVRANIIMLDDIGAEVTPWVRDEVIGPLHYRMVHELPTFFSSNFDYSELEHHA 129  
 DB 135 FLKTVANSFVLDVDDGSRSLSDWR-ELISYITITRYNNKSLTITTY-----S 184  
 QY 130 MTRDGEKTK-----AARIIEV-----KSLSTP 153  
 DB 185 LQREBSVRISSADLASRLGENVSKTYEMNELVTKSGDLRKSKLSTP 234

## RESULT 13

D69945  
 phage-related protein homolog ydam - Bacillus subtilis  
 C/Species: Bacillus subtilis

C|Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999  
C|Accession: D69945  
A|Authors: F. Ogasawara, N. Moszer, I. Albertini, A.M. Alloni, G. Azevedo, V. Bertenz  
R.Kunst, P. Broillett, S. Brusch, C.V. Caldwell, B. Capiano, V. Carer, N.M. Ch  
C. |, Brion, S. Broillett, S. Brusch, C.V. Caldwell, B. Capiano, V. Carer, N.M. Ch  
A. |, Ehrlich, S.D. Emerson, P.T. Entian, K.D. Ewington, J. Fabret, C. Ferrati, E.  
Nature 390, 249-256, 1997  
A|Authors: Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A., Gallen  
lech, J., Harwood, C.R., Henauf, A., Hilbert, H., Holappel, S., Hosono, S., Hullo, M.F.  
Koetter, P., Konigstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Lardin  
A|Authors: Labber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H., Masuda, S., Maue  
Y. M., Ogawa, K., Ogiwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M., Port  
Rieger, M., Rivolta, C., Kocho, E., Kocho, B., Rose, M., Sadle, J., Sato, T., Scanon  
A|Authors: Schleif, S., Schroeder, R., Scofield, F., Sekiguchi, J., Seko, A., Serocin  
akeuchi, M., Tamaki, A., Tanaka, T., Terpetra, P., Tsognou, A., Tosato, V., Uchiyama  
T. |, Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yamamoto, K., Yate, K., Yoshida  
A|Authors: Yoshikawa, H.F., Zumbstein, E., Yoshikawa, H., Danchin, A.  
A|Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A|Reference number: A69580; MUID: 98044033; PMID: 9384377  
A|Accession: D69945  
A|Status: preliminary; nucleic acid sequence not shown; translation not shown  
A|Molecule type: DNA  
A|Residues: 1-313 <KDN>  
A|Cross-references: GB:Z99117; GB:AL009126; NID: 92634966; PIDN: CAB14567.1; PID: e1183855;  
A|Experimental source: strain 168  
C|Genetics:  
A|Gene: ygdM

Query March Similarity 15.0%; Score 127.5; DB 2; Length 313;  
Best Local Similarity 33.1%; Pred. No. 0.00076;  
Matches 40; Conservative 19; Mismatches 49; Indels 13; Gaps 6;

Dy 12 GEQVGLTVLPFGTGSFPIIGAIANOL-KSKKVRSTIIYLP--EPIITLKGCFKDSFE 68  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Dd 161 GERONSIALLGQPSGSKTHLLTALIMNNLIKKSVAH--CMFYPYVEGMGDLLXAF--DML 216  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Dy 69 KKLHRVZRNAILMDLI-----GAEEVPWVRDEYLGILLHYRWVHELPPTSSNFYSX 123  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Dd 217 AKDLAMRKVEVLFTIDLEPKPINGQPRATDW-QVEIOISLVNRYRILNHPELLISSSELTIDE 275  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Dy 124 L 124  
Dd 276 I 276

RESULT 14  
I40411  
PBsx prophage ORF xkdc - Bacillus subtilis  
N:Alternate names: protein 5 (xre region)  
C:Species: Bacillus subtilis  
C:Date: 12-Aug-1996 #sequence revision 12-Aug-1996 #text\_change 15-Oct-1999  
C:Accession: I40411; C69711; S4711  
J:McDonnell, G.E.; Wood, H.; Devine, K.M.; McConnell, D.J.  
J: Bacteriol. 176: 5820-5830, 1994  
A>Title: Genetic control of bacterial suicide: regulation of the induction of PBsx in B.

A:Reference number: I40408; MUID:94364963; PMID:8083174  
A:Accession: I40411  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-266 <RES>  
A:Cross-references: EMBL:Z34287; NID:g498610; PIDD:CA84044.1; PID:g498614  
A:Kunst, F.; Ogatawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azavedo, V.; Bertet,  
C.; Bron, S.; Brulliet, S.; Bruschl, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.; Chai,  
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallen,  
teich, U.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsspel, S.; Hosono, S.; Hullio, M.F.;  
Koester, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidis, A.; Lardinols,  
A:Authors: Lauber, U.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauee,  
Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle,  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,  
A:Authors: Schleib, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serovska,  
Akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Togononi, A.; Tosato, V.; Uchiyama,  
J.; Winiers, P.; Wipac, A.; Yamamoto, H.; Yanane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.

A;Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A;Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*  
A;Reference number: A69580; MUID:98044033; PMID:9384377  
A;Accession: C69731  
A;Stratus: nucleic acid sequence not shown; translation not shown  
A;Molecule\_type: DNA  
A;Residues: 1-338, 'K', 240-266 <KUN>  
A;Cross-references: GB:Z29110; GB:AL009126; NID:g2633472; PIDN:CAB3110.1; PID:e1183272  
A;Experimental source: strain 168  
C;GeneticB:  
A;Gene: xkdc

Query Match 14.9%; Score 127; DB 2; Length 266;  
 Best Local Similarity 31.0%; Pred. No. 0.00069;  
 Matches 39; Conservative 25; Mismatches 44; Indels 18; Gaps 7

Qy	6	CTA--ITNGBOVK-----GLYLYGPFPTGKSPILGALINOLSKKVRSTIIYLP--EPIR	56
Db	101	CTKREYVADYEDIKCRKNISIALTLQPGSGKTHLLTAAANEL--MRCYVPVYIFPVEGPT	159
Qy	57	TLKGGFGDGSSEKQLHVRREANIIMLDDI-----GAEVLTWVRDEYVGPLLHYMAHEL	111
Db	160	DLKNDP--ALAEATLNRMKQADVIFIDDLFRPVGVKPRATDMOLEQMS--VLNRYEYLNHK	216
Qy	112	PTFFSS	117
Db	217	PILSS	222

RESULT 15  
D81415

**Chromosomal replication initiator protein Cj0001** [imported] - Campylobacter jejuni (strain ATCC 35061)  
C:Species: Campylobacter jejuni  
C>Date: 31-Mar-2000 #sequence\_revision: 31-Mar-2000 #ext\_change: 03-Jun-2002  
C/Acession: D81445  
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillingworth, T.P.; Davis, A.A.; Dougan, G.; Fraser, H.; Parkhill, J.; Quail, M.; Rajadaram, M.A.; Rutherford, K.M.; Vanyilet, A.; Whitehead, S.; Barrett, C.W.  
N:Nature 403, 665-668, 2000  
A>Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypoxanthine auxotrophy  
A/Reference number: AB1250; MUID:20150912; PMID:10688204  
A/Acession: D81445  
A>Status: preliminary  
A:Molecule type: DNA  
A/Residues: 1-440 <PAR>  
A/Cross-references: GB:A1139074; GB:A111168; NID:g6967505; PIDB:CAB72494.1; PIR:g6967500  
A/Experimental source: serotype O2, strain NCIC 11168  
C/Genetics:  
A/Gene: dnaA; Cj0001  
A/Superfamily: replication initiation protein dnaA

Query Match	14.7%	Score 125.5;	DB 2;	Length 440;
Best Local Similarity	26.8%	Pred. No. 0.0017;		
Matches 41; Conservative	30;	Mismatches 51;	Indels 31;	Gaps 7

```

QY 6 CTAITNGEOWKGLY-----LYGPEFGKSPILGAINOLKSKKVRSTIYLL--PEPIRTLK 59
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Db 123 CKAIHKDKGKTKIYNDIFVYGPPTGKTHLLDAQVN--ASLEMGKVIYANSNFINDPT 18
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
QY 60 GGFQDGSPEKKLRHVEANIIMLDDIGAEVYPPWRDEVYGLPLHYRMVHELPPFSNF 11
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Db 181 SNLKNSLDDFKHEKRYKNCVLLIDVQPLGKTDKIQEE-----FF--F 22
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
QY 120 DYSELBHH---LMTRDGEKTKAARIYERVKS 149
      :: : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Db 222 IFNEIKNDQIIMTSNDPNPMLKG--IIRRLKS 253
      :: : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||

```

Search completed: December 15, 2003, 15:16:49  
Job time : 9.24897 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 15:11:40 ; Search time 5.1845 Seconds  
(without alignments)  
1486.453 Million cell updates/sec

Title: US-09-689-952-16

Perfect score: 851  
Sequence: 1 AADICTAITNGEQVKGVL.....ERVKSLSTPPFLSGENFRNN 164

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41.\*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	427.5	50.2	311 1	DNAI_BACSU
2	127.5	15.0	313 1	YOMI_BACSU
3	127	14.9	266 1	XKDC_BACSU
4	125.5	14.7	440 1	DNA_CAMJE
5	114	13.4	265 1	ISTB_PSBAS
6	106.5	12.5	246 1	DNA_BUCAP
7	106	12.5	457 1	DNA_BUCBP
8	105.5	12.4	229 1	ISTB_BURCE
9	104.5	12.3	248 1	YDAV_ECOLI
10	103.5	12.2	263 1	Y4BW_RHISN
11	103	12.1	454 1	DNA_BUCAI
12	102.5	12.0	246 1	DNA_BUCAI
13	102.5	12.0	432 1	SKDI_SCHPO
14	101	11.9	251 1	ISTB_BACST
15	100	11.8	307 1	RUVB_MYCGE
16	100	11.8	462 1	DNA_YERPE
17	100	11.8	464 1	DNA_SERMA
18	100	11.8	643 1	DNA_STRRE
19	100	11.8	656 1	DNA_STRRO
20	99.5	11.7	231 1	ISTB_PSEFL
21	99	11.6	466 1	DNA_PROMI
22	98	11.5	454 1	DNA_BUCAP
23	98	11.5	466 1	DNA_SALTI
24	98	11.5	466 1	DNA_SALTY
25	98	11.5	467 1	DNA_ECO57
26	98	11.5	467 1	DNA_ECOLI
27	97.5	11.5	451 1	DNA_PASMU
28	97	11.4	245 1	DNA_ECOLI
29	97	11.4	263 1	ISTB_BACFR
30	96	11.3	410 1	PSMR_METTH
31	95.5	11.2	340 1	RFC5_HUMAN
32	95.5	11.2	453 1	DNA_STRPN
33	95	11.2	468 1	DNA_VIRBV

34	95	11.2	624 1	DNA_STRCH	09zh75 streptomyc
35	95	11.2	758 1	SC18_YEAST	18r759 saccharomyc
36	94.5	11.1	399 1	DNA_AQUAE	066559 aquifex aco
37	94.5	11.1	794 1	SC18_CANAL	P34732 candida alb
38	94	11.0	398 1	PSMR_ARCFU	028303 archaeoglob
39	94	11.0	436 1	PSMR_METKA	08x003 methanopyru
40	94	11.0	451 1	DNA_LISNO	092lv2 listeria in
41	94	11.0	451 1	DNA_LISNO	08yav2 listeria in
42	94	11.0	484 1	DNA_ZYMO	098493 zymomonas m
43	93	10.9	463 1	PS4_YEAST	P52917 saccharomyc
44	93	10.9	463 1	DNA_RICPR	059758 rickettsia
45	92	10.8	250 1	ISTB_BACTB	099338 bacillus th

## ALIGNMENTS

RESULT 1	ID	DNAI_BACSU	STANDARD;	PRT;	311 AA.
AC	P06567				
DT	01-JAN-1988	(Rel. 06, Created)			
DT	01-JAN-1988	(Rel. 06, Last sequence update)			
DT	28-FEB-2003	(Rel. 41, Last annotation update)			
DE	Primosomal protein dnaI.				
GN	DNAI				
OS	Bacillus subtilis.				
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.				
OX	NCBI_TaxID=1423;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=87117549; PubMed=3027671;				
RA	Ogasawara N., Moriya S., Mazza P.G., Yoshikawa H.;				
RT	"Nucleotide sequence and organization of dnaB gene and neighbouring				
RT	genes on the Bacillus subtilis chromosome."				
RL	Nucleic Acids Res. 14:9989-9999 (1986).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=97124191; PubMed=8969504;				
RA	Wipat A., Carter N., Brignell C.S., Guy J.B., Piper K.;				
RT	Sanders J., Emerson P.T., Harwood C.R.;				
RT	"The dnaB-phaA (256 degrees-240 degrees) region of the Bacillus				
RT	subtilis chromosome containing genes responsible for stress				
RT	responses, the utilization of plant cell walls and primary				
RL	metabolism."				
RN	Microbiology 142:3067-3078 (1996).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=168;				
RX	MEDLINE=98048467; PubMed=9387221;				
RA	Lapides A., Galleron N., Sorokin A., Ehrlich S.D.;				
RT	"Sequencing and functional annotation of the Bacillus subtilis genes				
RT	in the 200 kb rmb-dnaB region."				
RL	Microbiology 143:3431-3441 (1997).				
RN	[4]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=168;				
RX	MEDLINE=98044033; PubMed=9384377;				
RA	Kuntz F., Ogasawara N., Mosser I., Albertini A.M., Alloni G.;				
RT	Azeredo V., Bartero M.G., Bessieres P., Bolotin A., Borchert S.;				
RT	Borrius R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.;				
RA	Bouillier S., Brusch C.V., Caldwell I.B., Capuano V., Carter N.M.;				
RA	Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.;				
RA	Dentant K.F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.;				
RA	Eutenei Z., Ewington J., Fabret C., Ferrari B., Foulger D.;				
RA	Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.;				
RA	Ghim S.Y., Glaser P., Goffeau A., Colightly E.J., Grandi G.;				
RA	Gutseppe G., Guy B.J., Haga K., Haeck J., Harwood C.R., Henaut A.;				
RA	Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.;				
RA	Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.;				
RA	Kobayashi Y., Koester P., Konigsstein G., Krogh S., Kumano M.;				
RA	Kurita K., Lapides A., Lardinois S., Lauber J., Lazarevic V.;				

RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,  
RA Paro V., Pohl T.M., Portetelle D., Portwolk S., Prescott A.M.,  
RA Pressac E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,  
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
RA Sorokin A., Taccanti E., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Takemura K., Tanaka T., Terpstra P., Tognoni A.,  
RA Tosafo V., Uchiyama S., Vandendol M., Vannier F., Vassaretto A.,  
RA Viari A., Wambut R., Wedler E., Wedler H., Wetzenecker T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,  
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
RT *subtilis*.";  
RL Nature 390:249-256(1997).  
RP [5]  
RP SEQUENCE OF 1-206 FROM N.A.  
RP MEDLINE=87118226; PubMed=3027697;  
RA Hoshino T., McKenzie T., Schmidt S., Tanaka T., Sueoka N.,  
RT "Nucleotide sequence of *Bacillus subtilis* dnaB: a gene essential for  
RT DNA replication initiation and membrane attachment.";  
RL Proc. Natl. Acad. Sci. U.S.A. 84:653-657(1987).  
RP [6]  
RP SIMILARITY TO DNA.  
RA MEDLINE=92195821; PubMed=1549481;  
RA Koonin E.V.,  
RT "Archaeobacterial virus SSV1 encodes a putative DnaA-like protein.";  
RL Nucleic Acids Res. 20:1143-1143(1992).  
RP [7]  
RP IDENTIFICATION.  
RA MEDLINE=95291463; PubMed=7773414;  
RA Brand C., Ehrlich S.D.,  
RT "The *Bacillus subtilis* dnaI gene is part of the dnaB operon.";  
RL Microbiology 141:1199-1200(1995).  
CC -1- FUNCTION: PROBABLY INVOLVED IN DNA REPLICATION.  
CC -1- SIMILARITY: SOME, TO DNA FROM VARIOUS BACTERIA.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X04963; CAA28633.1; -;  
DR EMBL: 275208; CAA99605.1; -;  
DR EMBL: AF008220; AAC00359.1; -;  
DR EMBL: Z99118; CAB14858.1; -;  
DR EMBL: M15183; AAA22405.1; -;  
DR PIR: B24720; I0BS44.  
DR Subtilisin; Bg10359; dnaI.  
DR InterPro: IPR003593; AAA\_ATPase.  
DR SMART: SMO0382; AAA: 1  
KW Primosome; DNA replication; ATP-binding; Complete proteome.  
FT NP BIND 168 175 ATP (PROBABLY).  
FT FT 19 K -> N (IN REF. 5).  
FT FT 24 M -> T (IN REF. 5).  
FT FT 24 M -> T (IN REF. 5).  
SQ SEQUENCE 311 AA; 36114 MW; A86FC94AB6841264 CRC64;  
Query Match 50.2%; Score 427.5; DB 1; Length 311;  
Best Local Similarity 55.7%; Pred. No. 1.8e-31;  
Matches 83; Conservative 26; Mismatches 39; Indels 1; Gaps 1;  
QY 16 KGLVYGFPGKSPFIATANOLKSKVSTIYLEBFRITLKGFKDSPEKQHRVR 75  
DB 162 KGLVYGFPGKSPFIATANOLKSKVSTIYLEBFRITLKGFKDSPEKQHRVR 221  
QY 76 EANILMDIDIGAEVTPWVDEVIQPLIHYRMVHEPTFFSSNDEYSELEHILMTDGE 135  
DB 222 TTPVLMDDIGABESMTSWVDEVIQVIGLQHRMSGQLPTFFSSNSNPDELKHHFYSGRGE 281

QY 136 -EXTRAKTIERPKYSTPYFLSGENFRN 163  
DB 282 KESEVKARILMERILYLAAPRIIDGENRRH 310  
RESULT 2  
ID YQAM\_BACSU STANDARD, PRT; 313 AA.  
AC P45910;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DB Hypothetical protein yqam.  
GN YQAM.  
OS *Bacillus subtilis*.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RP [1]  
RP SEQUENCE FROM N.A.  
RP STRAIN=168 / JH642;  
RA MEDLINE=95219086; PubMed=7704261;  
RA Takemaru K.-I., Mizuno M., Sato T., Takeuchi M., Kobayashi Y.,  
RT "Complete nucleotide sequence of a skin element excised by DNA  
RT rearrangement during sporulation in *Bacillus subtilis*.";  
RL Microbiology 141:323-327(1995).  
RP [2]  
RP SEQUENCE FROM N.A.  
RP STRAIN=168 / JH642;  
RA MEDLINE=97124195; PubMed=8969508;  
RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,  
RA Kobayashi Y.,  
RT "Systematic sequencing of the 283 kb 210 degrees-223 degrees region of  
RT the *Bacillus subtilis* genome containing the skin element and many  
RT sporulation genes.";  
RL Microbiology 142:3103-3111(1996).  
RP [3]  
RP SEQUENCE FROM N.A.  
RP STRAIN=168;  
RA MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogatawara N., Moser I., Albertini A.M., Alloni G.,  
RA Azevedo V., Betero M.G., Bessieres P., Bolocin A., Borchert S.,  
RA Borriss R., Bourcier L., Brans A., Braun M., Brignell S.C., Bron S.,  
RA Brouillet S., Bruch C.V., Caldwell I.F., Cummings N.J., Daniel R.A.,  
RA Choi S.K., Codani J.U., Comercon I.F., Cummings N.J., Daniel R.A.,  
RA Dautot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
RA Denton K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
RA Ghim S.Y., Glaeser P., Goffeau A., Golligly E.J., Grandi G.,  
RA Giuseppe G., Guy B.J., Haga K., Haelech J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kasahara Y., Klaert-Blanchard M., Klein C.,  
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
RA Kuriyama K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,  
RA Paro V., Pohl T.M., Portetelle D., Portwolk S., Prescott A.M.,  
RA Pressac E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,  
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
RA Sorokin A., Taccanti E., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Takemura K., Tanaka T., Terpstra P., Tognoni A.,  
RA Tosafo V., Uchiyama S., Vandendol M., Vannier F., Vassaretto A.,  
RA Viari A., Wambut R., Wedler E., Wedler H., Wetzenecker T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,  
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
RT *subtilis*.";  
RL Nature 390:249-256(1997).  
RP [4]  
RP IDENTIFICATION.

RA MEDLINE=96084975; PubMed=7489895;  
 RA Medigue C., Moszer I., Viari A., Danchin A.;  
 RA "Analysis of a Bacillus subtilis genome fragment using a co-operative  
 RT computer system prototype";  
 RL Gene 165:GC37-CC51(1995).  
 CC -1- SIMILARITY: TO B.SUBTILIS Y0XC AND T.HYDROXYBENTYL HEMOLYSIN  
 TLVA.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; D32216; BAA06927.1; -  
 DR EMBL; D84432; BAA12388.1; -  
 DR EMBL; Z99117; CAB14567.1; -  
 DR PIR; D69945; D69945.  
 DR Subtilisin; BG11264; ygaM.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR SMART; SM00382; AAA; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 313 AA; 36138 MW; 09208795E310D73A CRC64;  
 Query Match 15.0%; Score 127.5; DB 1; Length 313;  
 Best Local Similarity 33.1%; Pred. No. 0.00023;  
 Matches 40; Conservative 19; Mismatches 49; Indels 13; Gaps 6;  
 QY 12 GEQVKGILYVGPFGTGSFTLGAIANQL-KSKKYRSTIYLP--EFIRLKGFGKGSPE 68  
 DB 161 GERGNISALLGQSGSKTHLTLAANLKKSVH--CMFVPEVGEKGLKAF--DNLE 216  
 QY 69 KKLHVRVANIIMLDI-----GAEVTPVRDVGIPLHYRMVHELPTFFSSNDYSE 123  
 DB 217 AKLDARKEVLFIDLFKPINQPRATDW-QVEQLQSVLVNRYLVNHNKPLISSSELTIDE 275  
 QY 124 L 124  
 DB 276 I 276  
 RESULT 3  
 XKDC\_BACSU STANDARD; PRT; 266 AA.  
 ID XKDC\_BACSU STANDARD; PRT; 266 AA.  
 AC P39782;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Phage-like element PBSX protein XKDC.  
 GN XKDC.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OC NCBI\_TaxID=1423;  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / SO113;  
 RC MEDLINE=94364963; PubMed=8083174;  
 RA McDonnell G.E., Wood H., Devine K.M., McConnell D.J.;  
 RT "Genetic control of bacterial suicide: regulation of the induction of  
 RT PBSX in Bacillus subtilis";  
 RT J. Bacteriol. 176:5820-5830(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RC Krogh S., O'Reilly M., Nolan N., Devine K.M.;  
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RC MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,

RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Boris R., Bourcier L., Brans A., Braun M., Britnell S.C., Bron S.,  
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,  
 RA Dentat F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Entian K.D., Errington J., Fabret C., Ferrati E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly B.J., Grandi G.,  
 RA Giuseppe G., Guy B.J., Haga K., Haeche J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapdus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestel D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,  
 RA Paro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Prescan E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadate Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone P.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Taccioni E., Takagi Y., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandendol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Welzenegger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,  
 RT The complete genome sequence of the Gram-positive bacterium Bacillus  
 RT subtilis";  
 RL Nature 390:249-256(1997).  
 CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTIONAL ANTITERMINATOR.  
 CC -1- SIMILARITY: STRONG, TO B.SUBTILIS Y0XM.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; Z34287; CAA84044.1; -  
 DR EMBL; Z70177; CAA94054.1; -  
 DR EMBL; Z99110; CAB13110.1; -  
 DR PIR; I40411; I40411.  
 DR Subtilisin; BG10996; xkdc.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR SMART; SM00382; AAA; 1.  
 KW ATP-binding; Complete proteome.  
 FT NP BIND 124 131  
 FT CONFIDENCE 239 239 K -> R (IN REF. 1).  
 SQ SEQUENCE 266 AA; 30534 MW; 9929C991E9D655AA CRC64;  
 Query Match 14.9%; Score 127; DB 1; Length 266;  
 Best Local Similarity 31.0%; Pred. No. 0.00021;  
 Matches 39; Conservative 25; Mismatches 44; Indels 18; Gaps 7;  
 QY 6 CTA--ITNGEYK-----GLVYVGPFGTGSFTLGAIANQLSKKYRSTIYLP--EFIR 56  
 DB 101 CTKEVADYDEQIKDKRKNSIALLGQSGSKTHLTPAANL-MRTCYVVIYVFPFEGFT 159  
 QY 57 TLKGGKGSFKKLHVRVANIIMLDI-----GAEVTPVRDVGIPLHYRMVHEL 111  
 DB 160 DLKNDP--ALTEAKLNRMKQADVLTIDLFKRVNGKPRATDWQLEQWYS-VLNYRYLNHK 216  
 QY 112 PTFSS 117  
 DB 217 PILLS 222  
 RESULT 4  
 DNAA\_CAMJ STANDARD; PRT; 440 AA.  
 ID DNAA\_CAMJ STANDARD; PRT; 440 AA.

AC 09PJ001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Chromosomal replication initiator protein dnaA.  
 GN DnaA OR C30001.  
 OS Campylobacter jejuni.  
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
 OC Campylobacteraceae; Campylobacter.  
 OK NCBI\_TaxID=197;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCTC 11168;  
 RX MEDLINE=20150912; PubMed=10688204;  
 RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,  
 RA Basham D., Chillingworth T., Davies R.M., Feltham T., Holtroyd S.,  
 RA Jaseis K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,  
 RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,  
 RA Whitehead S., Barrall B.G.;  
 RA The genome sequence of the food-borne pathogen Campylobacter jejuni  
 RT reveals hypervariable sequences.";  
 RL Nature 403:665-668(2000).  
 CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE INITIATION AND REGULATION  
 CC OF CHROMOSOMAL REPLICATION. BINDS TO THE ORIGIN OF REPLICATION;  
 CC (DNA BOX): 5'-TTATC(C/A)(C/A)A-3'. DnaA BINDS TO ATP AND TO  
 CC ACIDIC PHOSPHOLIPIDS (BY SIMILARITY).  
 CC -1- SIMILARITY: Belongs to the dnaA family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AL139074; CAB72494.1; -  
 DR PIR: D81415; D81415.  
 DR HAMAP: MF\_003777; -; 1.  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR InterPro: IPR001957; Bac\_DnaA.  
 DR Pfam: PF00308; bac\_dnaA; 1.  
 DR PRINTS: PR00051; DnaA.  
 DR SMART: SM00382; AAA; 1.  
 DR TIGRFAMs: TIGR00362; DnaA; 1.  
 DR PROSITE: PS01008; DnaA; FALSE\_NEG.  
 KW DnaA replication; DNA-binding; ATP-binding; Complete proteome.  
 FT NP BIND 143 150 ATP (POTENTIAL).  
 SQ SEQUENCE 440 AA; 49686 MW; A3FAB90CCED06BD7 CRC64;  
 Query Match 14.7%; Score 125.5; DB 1; Length 440;  
 Best Local Similarity 26.8%; Pred. No. 0.00052;  
 Matches 41; Conservative 30; Mismatches 51; Indels 31; Gaps 7;  
 QY 6 CTATNTEGVKGY----LYGPGTGSFLLGAIANDLKSXYRSTIYV--EPIRTLK 59  
 DB 123 CKAIAHDKKQKGINPIYVGTGLGKTHLQAVGN--ASLENGKVIYATSENFINDFT 180  
 QY 60 GGFKDGSEFKKQIHRVREANITMLDDIGAEVTPWVRDEVIGPLLYHVMVHELPFFSSNF 119  
 DB 181 SNKNGSLDGFHKKYRNCVDLLIDVQFLGKTIDKIQER-----FF---F 221  
 QY 120 DYSELEHH--LMTRDGEKTKAARIIEKVS 149  
 DB 222 IFNEIKNDQIIMTSDNPPMLKG-ITERLKS 253  
 RESULT 5  
 ID ISTB\_PSEAE STANDARD; PRT; 265 AA.  
 AC P15026;  
 DT 01-APR-1990 (Rel. 14, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Insertion sequence IS21 putative ATP-binding protein.  
 GN ISTB.  
 OS Pseudomonas aeruginosa.  
 OS Plasmid R68.45.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 OK NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=89218951; PubMed=2540414;  
 RA Reimann C., Moore R., Little S., Savioz A., Willetts N.S., Haas D.;  
 RT "Genetic structure, function and regulation of the transposable  
 RT element IS21.";  
 RL Mol. Gen. Genet. 215:416-424(1989).  
 RN [2]  
 RP REVISION TO 283.  
 RA Berger B.;  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.  
 CC -1- FUNCTION: THE ISTB PROTEIN, ONE OF TWO PROTEINS EXPRESSED ONLY  
 CC WHEN THERE IS A TANDDEM REPEAT OF THE IS21 INSERTION SEQUENCE, IS  
 CC NECESSARY FOR THE TRANSPOSITION OF PLASMIDS WITH THAT TANDDEM  
 CC REPEAT.  
 CC -1- SIMILARITY: BELONGS TO THE IS21/IS1162 PUTATIVE ATP-BINDING  
 CC PROTEIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X14793; CAA32899.2; -  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR InterPro: IPR002611; IstB\_ATPbind.  
 DR Pfam: PF01695; IstB; 1.  
 DR SMART: SM00382; AAA; 1.  
 KW Plasmid; Transposable element; ATP-binding.  
 FT NP BIND 104 111 ATP (POTENTIAL).  
 SQ SEQUENCE 265 AA; 30528 MW; 904C7ADC5B3A12A9 CRC64;  
 Query Match 13.4%; Score 114; DB 1; Length 265;  
 Best Local Similarity 26.1%; Pred. No. 0.0032;  
 Matches 42; Conservative 36; Mismatches 53; Indels 30; Gaps 8;  
 QY 13 EGVKGLYVPGTGSFLLGAIANDLKSXYRSTIYV--EPIRTLKGFPGDSFEKK 70  
 DB 95 ERSENVILGPPGVGKTHL--AIALGVKAVDAGHRVLPDLRLITLTKAKQENRLERQ 152  
 QY 71 LHRVREANITMLDDIGAEVTPWVRDEVIGPLLYHVMVHELPFFSSNF--FDYSE 123  
 DB 153 LQOLSVARVILIDELG--YLPNNRREA--SLPFLRLRRYRKASIIILTSNGKGFADMB 206  
 QY 124 L--EHLMTNRDGEKTKAARIIEKVSSTYPLSGENFR 162  
 DB 207 MFGDHYL-----ATAIDRLHLHSTLTINKGESYR 236  
 RESULT 6  
 ID DNAC\_BUCAP STANDARD; PRT; 246 AA.  
 AC O8KA79;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE DNA replication protein dnaC.  
 GN DNAC OR BUSG022.  
 OS Buchnera aphidicola (subsp. Schizaphis graminum).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Buchnera.

OX NCBI\_TaxID=98794;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22084549; PubMed=12089438;  
 RA Tamas I., Klasson L., Candaeck B., Naeslund A.K., Eriksson A.-S.,  
 Wernegren U.J., Sandstrom J.P., Moran N.A., Andersson S.G.E.;  
 RT "50 million years of genomic stasis in endosymbiotic bacteria.";  
 RL Science 296:2376-2379(2002).  
 CC -1- FUNCTION: THIS PROTEIN IS REQUIRED FOR CHROMOSOMAL REPLICATION. IT  
 FORMS, IN CONCERT WITH DNAB PROTEIN AND OTHER PREPIMING PROTEINS  
 DNAT, N', N', A PREPIMING PROTEIN COMPLEX ON THE SPECIFIC SITE  
 OF THE TEMPLATE DNA RECOGNIZED BY PROTEIN N' (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE DNAC FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; AE014077; AAM67594.1;  
 KW DNA replication; Primosome; Complete proteome.  
 FT SITE 69  
 FT SEQUENCE 246 AA; 28497 MW; BF7E2A9BC2ADB5D CRC64;  
 SQ  
 Query Match 12.5%; Score 106.5; DB 1; Length 246;  
 Best Local Similarity 25.2%; Pred. No. 0.014;  
 Matches 35; Conservative 29; Mismatches 60; Indels 15; Gaps 4;  
 OY 13 EGVKGLYLKPGRTGKSPFLGATANOLKSKKRVSTIYLPERRTLKGFPGKSGS---PRK 69  
 Db 97 ENIASFTSGRPGTGNHLASALGNLILHGKSLILVTAADLMSNKGFGSGTSNTEEN 156  
 OY 70 KLRVREANILMDLDIGAEVTPWVDEVIQPLHYRMVHELPTFPSSNFDYSELEHHA 129  
 Db 157 LHMNLSVDLIMDELGMGTSEYER-VLIINQVDRSSSKSTGMLSLNDRGMGNLL- 214  
 OY 130 MTRDGEKTKAARIIRVK 148  
 Db 215 ---GE-----RVIDRMR 223  
 RESULT 7  
 DNAA\_BUCBP  
 ID DNAA\_BUCBP STANDARD; PRT; 457 AA.  
 AC P59567;  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Chromosomal replication initiator protein dnnaA.  
 GN DNAA OR BBP012.  
 OS Buchnera aphidicola (subsp. Baitongia pistaciace).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Buchnera.  
 NCBI\_TaxID=135842;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426901; PubMed=12522265;  
 RA Van Ham R.C.H.J., Kamerbeek J., Palacios C., Rausell C., Abascal F.,  
 Bascolla U., Fernandez J.M., Jimenez L., Postigo M., Silva F.J.,  
 Tames J., Viguera B., Latorre A., Valencia A., Moran F., Moya A.;  
 RT "Reductive genome evolution in Buchnera aphidicola.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:581-586(2003).  
 CC -1- FUNCTION: Plays an important role in the initiation and regulation  
 of chromosomal replication. Binds to the origin of replication; it  
 binds specifically double-stranded DNA at a 9 bp consensus (dnna  
 box): 5'-TATC(A/A)A/C(A/A)-3'. DnaA binds to ATP and to acidic  
 phospholipids (By similarity).  
 CC -1- SIMILARITY: Belongs to the dnna family.  
 CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; AE014016; AAO26756.1;  
 DR HAMAP; MF\_00377; -; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS01008; DNAA; FALSE NEG.  
 KW DNA replication; DNA-binding; ATP-binding; Complete proteome.  
 FT NP BIND 162  
 FT SEQUENCE 457 AA; 53365 MW; B2950813FADDA028 CRC64;  
 SQ  
 Query Match 12.5%; Score 106; DB 1; Length 457;  
 Best Local Similarity 24.1%; Pred. No. 0.032;  
 Matches 41; Conservative 30; Mismatches 71; Indels 28; Gaps 6;  
 OY 18 LYLVPFGVKSFLIGATANOLKSKKRVSTIYLPERRTLKGFPGDGSFEKLRVR 75  
 Db 158 LFLYKSGGLGKTHLHAAVANTLKYKNTIKIYINSENFQMTSLKNTLIEFKKYR 217  
 OY 76 EAMTMDLDIGAEVTPWVDEV---IGPLHY-----RMVHELPTFPSSNPD 120  
 Db 218 SVNTLIDIDIPRAVKSGSELEFHTNALNRNOQIITTSQPFQKHGIETRLKSR- 276  
 OY 121 YSELEHHLAMTRDGEKTKAARI---ERVKSLSTPY---FLSGENFRNN 164  
 Db 277 ---BCGLTIRIDPPDLMTRTKILIKSHIYDINLSKYVAFIANKLKN 322  
 RESULT 8  
 ISTB\_BURCE  
 ID ISTB\_BURCE STANDARD; PRT; 229 AA.  
 AC P55923;  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Insertion sequence IS401, a new member of the IS3 family implicated  
 in plasmid rearrangements in Pseudomonas cepacia.";  
 OS Burkholderia cepacia (Pseudomonas cepacia).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Burkholderiaceae; Burkholderia.  
 NCBI\_TaxID=292;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=ATCC 17616 / 249;  
 RX MEDLINE=94302134; PubMed=7518087;  
 RA Byrne A.M., Leslie T.G.;  
 RT "Characteristics of IS401, a new member of the IS3 family implicated  
 in plasmid rearrangements in Pseudomonas cepacia.";  
 RL Plasmid 31:138-147(1994).  
 CC -1- SIMILARITY: BELONGS TO THE IS21/IS1162 PUTATIVE ATP-BINDING  
 PROTEIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; L09108; -; NOT ANNOTATED CDS.  
 DR InterPro; IPR003593; AAA ATPase.  
 DR InterPro; IPR002611; IserB\_Atpbind.  
 DR Pfam; PF01695; IserB; 1.  
 DR SMART; SM00382; AAA; 1.  
 KW Transposable element; ATP-binding.  
 FT NP BIND 105  
 FT SEQUENCE 229 AA; 25929 MW; ACS619C729163E1 CRC64;  
 SQ

Query Match 12.4%; Score 105.5; DB 1; Length 229;  
 Best Local Similarity 26.8%; Pred. No. 0.016;  
 Matches 38; Conservative 29; Mismatches 64; Indels 11; Gaps 5;

QY 6 CTAITNGEYQVKGVLVYPFGTGSFIIIGAIANQLSKKVRSTIIYLPEFIRTLKGGF 65  
 DB 92 CDWRNQAQ---NLITGPTGAGKWLACAFQOACROGFSVFYVARLPEELRIAGDG 148  
 QY 66 SPEKKLHVRANILMDDIGAEVTPWVRDEVIGPLIHYRMHLELPFFSSNDYSELE 125  
 DB 149 SFTRRLQLAKIDVLLIDDLGOLDOAARNDL-EVLDDR-VGTRSTVITSQ---PLE 203  
 QY 126 HHLAMTRDGEKTKAARIERY 147  
 DB 204 HHHAWLQD---PTLADAILDLRL 222

RESULT 9  
 YDAB\_ECOLI STANDARD; PRT; 248 AA.

AC P7546;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DE Hypothetical protein ydab.  
 GN YDAB OR B1360.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=9725157; PubMed=9097039;  
 RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isano K., Itoh T.,  
 RA Makino H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,  
 RA Nakano K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
 RA Nakada S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Salto N.,  
 RA Sempel G., Seki Y., Sivasubraman S., Tagami H., Takeda J.,  
 RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horuchi T.;  
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome  
 corresponding to the 28.0-40.1 min region on the linkage map.";  
 RL DNA Res. 3:363-377(1996).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AE000233; AAC74442.1; -  
 EMBL; D90774; BAA1958.1; -  
 DR EIR; C64886; C64886.  
 DR EcoGene; EG1364; ydab.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR002611; IctB\_ATPbind.  
 DR Pfam; PF01695; IctB\_1.  
 DR SMART; SM00382; AAA; 1.  
 KW Hypothetical protein; Coiled coil; Complete proteome.  
 FT DOMAIN 33 57 COILED COIL (POTENTIAL).  
 SQ SEQUENCE 248 AA; 28128 MW; 3501DCB1E8512E01 CRC64;

Query Match 12.3%; Score 104.5; DB 1; Length 248;  
 Best Local Similarity 27.2%; Pred. No. 0.021;  
 Matches 40; Conservative 28; Mismatches 60; Indels 19; Gaps 6;

QY 2 ADDICTAITNGEYQVKGVLVYPFGTGSFIIIGAIANQLSKKVRSTIIYLPEFIRTLKGG 61  
 DB 94 ADELMTGCTN-----FAFGKRGPTGKGNHAAIIGRLKDGQTVVTVYADVMSALHAS 147  
 QY 62 FKQG-SPEKKLHVRANILMDDIGAEVTPWVRDEVIGPLIHYRMHLELPFFSSNFD 120  
 DB 148 YDDGSGEKFRLCEVDLVLDLDEIGIQRGT---KNEQV--VLH-QIVRRRTASMSVGM 201  
 QY 121 YSELHHLAMTRDGEKTKAARIERY 147  
 DB 202 LTVLNTYAMKTLIGE-----RIMDRN 222

RESULT 10  
 Y4BM\_RHISN STANDARD; PRT; 263 AA.

AC P5380;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE Putative insertion sequence ATP-binding protein Y4BM/Y4KI/Y4TA.  
 GN Y4BM AND Y4KI AND Y4TA.  
 OS Rhizodium sp. (strain NGR234).  
 OC Rhizobium; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.  
 OX NCBI\_TaxID=394;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97305956; PubMed=9163424;  
 RA Feilberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,  
 RA Perret X.;  
 RT "Molecular basis of symbiosis between Rhizobium and legumes.";  
 RL Nature 387:394-401(1997).  
 CC -1- SIMILARITY: BELONGS TO THE IS21/IS1162 PUTATIVE ATP-BINDING  
 CC PROTEIN FAMILY.  
 CC -1- SIMILARITY: STRONG, TO Y4UH, ALSO TO Y4IO/Y4ND/Y4SD AND Y4PL.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AE000066; AAB91628.1; -  
 EMBL; AE000081; AAB91740.1; -  
 DR EMBL; AE000097; AAB91855.1; -  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR002611; IctB\_ATPbind.  
 DR Pfam; PF01695; IctB\_1.  
 DR SMART; SM00382; AAA; 1.  
 KW Hypothetical protein; Plasmid; Transposable element; ATP-binding.  
 FT NP\_BIND 112 119 ATP (POTENTIAL).  
 SQ SEQUENCE 263 AA; 29607 MW; 756836C342DC955E CRC64;

Query Match 12.2%; Score 103.5; DB 1; Length 263;  
 Best Local Similarity 21.3%; Pred. No. 0.028;  
 Matches 37; Conservative 29; Mismatches 55; Indels 53; Gaps 7;

QY 8 AITNGEYQV---KGLVYYPFGTGSFIIIGAIANQLSKKVRSTIIYLPEFIRTLKGGF 62  
 DB 95 ALAAGQWLRNHLRLITGHTGKSWLACAFGQ--AARLGHVLYVRVPRMREELALAR 152  
 QY 63 KQSGFEKKLHVRANILMDDIG-----AEE-----VTPW 93  
 DB 153 LDGSPRLIDRLRVLQLLIDDLGTTLSDQGRFHLFEIYERYQRKSTLITAVVAVSW 212

QY 94 ---VRDEIGPLHYRWHELPFFSSNPDYSELEHILANTROGEKTKARII 144  
 DB 213 HDLADSTVADALIDRIVEN-----AHRITLRGSMKQKSAPIIL 252

## RESULT 11

DNA\_BUCAL STANDARD; PRT; 454 AA.

AC 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Chromosomal replication initiator protein dnaa.  
 GN DNA OR BU012

OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum  
 symbiotic bacterium).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Buchnera.  
 OX NCBI\_TaxId=118099;

RP SEQUENCE FROM N.A.  
 RC STRAIN=Tokyo 1998;

RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;  
 RT "Genome sequence of the endocellular bacterial symbiont of aphids  
 Buchnera sp. APS.";

RL Nature 407:81-86(2000).

CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE INITIATION AND REGULATION  
 OF CHROMOSOMAL REPLICATION. BINDS TO THE ORIGIN OF REPLICATION;  
 IT BINDS SPECIFICALLY DOUBLE-STRANDED DNA AT A 9 BP CONSENSUS  
 (DNA BOX): 5'-TTATC(C/A)(C/A)-3'. DNA BINDS TO ATP AND TO  
 ACIDIC PHOSPHOLIPIDS.

CC -1- SIMILARITY: Belongs to the dnaa family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AP001118; BAB12740.1; -

DR HAMAP; MF\_00377; -1.

DR InterPro; IPR003593; AAA\_ATPase.

DR Pfam; PF00308; bac\_dnaa; 1.

DR PRINTS; PRO0051; DNAA.

DR SMART; SM00382; AAA; 1.

DR TIGRFAMs; TIGR00362; Dnaa; 1.

DR PROSITE; PS01008; DNAA; 1.

KW DNA replication; DNA-binding; ATP-binding; Complete proteome.

FT NP BIND 159 166 ATP (POTENTIAL).

SO SEQUENCE 454 AA; 52931 MW; 1F5299F4C2751213 CRC64;

Query Match 12.1%; Score 103; DB 1; Length 454;  
 Best Local Similarity 32.9%; Pred. No. 0.058;  
 Matches 25; Conservative 15; Mismatches 34; Indels 2; Gaps 1;

QY 12 GEQVKGILYLPFGTGSFLIGAIANOLKSKVRSTIIVPE--PIRTLKGFPGDSFEK 69  
 DB 149 GNSYINPLFLGAGLGLGTHLHAIHGNILSYKIDIKITFMNSSECFVDMVKALKNNAIEK 208

QY 70 KLRVREANILMLDDI 85

DB 209 FKLTYRSVDALLDDI 224

RESULT 12  
 DNAC\_BUCAL STANDARD; PRT; 246 AA.  
 AC P57134;

DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update) -  
 DE DNA replication protein dnaa.  
 GN DNAC OR BU021.

OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum  
 symbiotic bacterium).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Buchnera.  
 OX NCBI\_TaxId=118099;

RP SEQUENCE FROM N.A.  
 RC STRAIN=Tokyo 1998;

RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;

RT "Genome sequence of the endocellular bacterial symbiont of aphids  
 Buchnera sp. APS.";

RL Nature 407:81-86(2000).

CC -1- FUNCTION: THIS PROTEIN IS REQUIRED FOR CHROMOSOMAL REPLICATION. IT  
 FORMS, IN CONCERT WITH DNA PROTEIN AND OTHER PREPRIMING PROTEINS  
 DNAT, N', N'', A PREPRIMING PROTEIN COMPLEX ON THE SPECIFIC SITE  
 OF THE TEMPLATE DNA RECOGNIZED BY PROTEIN N' (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE DNAC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AP001118; BAB12748.1; -

DR DNA replication; Primosome; Complete proteome.

FT SITE 69 PROBABLY INVOLVED IN THE INTERACTION  
 WITH THE DNAB PROTEIN (BY SIMILARITY).

SO SEQUENCE 246 AA; 28444 MW; C2E56A3BCBE1A68F CRC64;

Query Match 12.0%; Score 102.5; DB 1; Length 246;  
 Best Local Similarity 24.5%; Pred. No. 0.032;  
 Matches 34; Conservative 29; Mismatches 61; Indels 15; Gaps 4;

QY 13 EOYKGLYLPFGTGSFLIGAIANOLKSKVRSTIIVPEPIRTLKGFPGDS---FEK 69  
 DB 97 ENIASFTFSKPGTGNHLSAIGNVLILHGSILLYTADLMSNMKGFSGSNTTEEN 156

QY 70 KLRVREANILMLDDIGABEVTWPVDEVIGPLHYRWHELPFFSSNPDYSELEHIL 129  
 DB 157 LKHDSSVDLIMDEIGMTESRYEK-VIINOIVDRSSSKSTGMLSNLDHGMKSL- 214

QY 130 MTRDGEKTKAARIIEVK 148

DB 215 ---GE-----RVIDRMK 223

## RESULT 13

SKOL\_SCHPO STANDARD; PRT; 432 AA.

AC 009803;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Suppressor protein of bent/beds double mutants.

GN SPAC2G11.06.

OS Schizosaccharomyces pombe (fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomycetes.

OX NCBI\_TaxId=4896;

RN [1]

RP SEQUENCE FROM N.A.

RA Valencik M.L., Pringle J.R.;

Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

```

Db      337  IVADDAIMEPRKRIHT 352

RESULT 14
ID      1STB_BACST      STANDARD;      PRT;      251 AA.
AC      045619;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      01-NOV-1997 (Rel. 35, Last annotation update)
DE      Insertion sequence IS5376 putative ATP-binding protein.
OS      Bacillus stearothermophilus.
OC      Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.
CX      NCBI_TaxID=1422;
OK      [1]
RN      SEQUENCE FROM N.A.
RP      STRAIN=CU21;
RC      MEDLINE=93161466; PubMed=6382625;
RX      Xu K., He Z.-Q., Mao Y.-M., Shen R.-Q., Sheng Z.-J.;
RA      "On two transposable elements from Bacillus stearothermophilus.";
RL      Plasmid 29:1-9(1993).
CC      -1- SIMILARITY: BELONGS TO THE IS21/IS1162 PUTATIVE ATP-BINDING
CC      PROTEIN FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on ways
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.ebi.ch/announce/
CC      or send an email to license@ebi.ebi.ch).
CC      -----
CC      EMBL; X67861; CAA48046.1; -.
CC      PIR; S23889; S23889.
CC      InterPro; IPR003593; AAA_ATPase.
CC      InterPro; IPR002611; IctB_ATPbind.
DR      Pfam; PF01695; IctB_1.
DR      SMART; SM00382; AAA; 1.
KW      Transposable element; ATP-binding.
FW      NP BIND      105      112      ATP (POTENTIAL).
SQ      SEQUENCE      251 AA; 29286 MW; FA8F4F55B2FB527 CRC64;

Query Match      11.9%; Score 101; DB 1; Length 251;
Best Local Similarity 20.8%; Pred. NO. 0.045;
Matches 33; Conservative 35; Mismatches 65; Indels 26; Gaps 4;

QY      13  EGVGKLVGPFGPGKSPILGAIYANQKSKKVRSTIYLPDEFIRTKGFGFDSFEKKLH 72
DB      96  DRKRNITFLPGPGIGKTHLAISGEMAIARGYKTYFITADLVNQLRADQEGLEKKLR 155
QY      73  RVEANITMLMDLIGAESEVTPWVRDEVGPLHY-----RMVHELPTFFSSNPDSYSELH 126
DB      156  VFVGPETVLIIIDMGYKLDP-----NSAHYLFQYIARAYEHAPFILSNKSPGEM-- 205
QY      127  HLAITRDGE--EKTARIIERYKSIISTPYFSGENFR 162
DB      206  -----GEIVGDSVLTATAMLDRLIHHSITFNLKGESYR 237

RESULT 15
ID      1STB_BACST      STANDARD;      PRT;      307 AA.
AC      049425;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Holliday junction DNA helicase ruvb.
OS      RUVB OR MG359.
OC      Mycoplasma genitalium.
OC      Bacteria; Firmicutes; Mollicutes; Mycoplasmales; Mycoplasma.
CX      NCBI_TaxID=20977;
OK      [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Doughty T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bort K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
CC -1- FUNCTION: THE RUVA-RUVB COMPLEX IN THE PRESENCE OF ATP RENATURES
CC CRUCIFORM STRUCTURE IN SUPERCOILED DNA WITH PALINDROMIC SEQUENCE,
CC INDICATING THAT IT MAY PROMOTE STRAND EXCHANGE REACTIONS IN
CC HOMOLOGOUS RECOMBINATION. RUVAB IS AN HELICASE THAT MEDIATES THE
CC HOLODAY JUNCTION MIGRATION BY LOCALIZED DENATURATION AND
CC REANNEALING (BY SIMILARITY).
CC -1- SUBUNIT: FORMS A COMPLEX WITH RUVA.
CC -1- SIMILARITY: BELONGS TO THE RUVB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U39717; AAC71584.1; -.
DR PIR; G64239; G64239.
DR TIGR; MG359; -.
DR HAMAP; MF_00016; -.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003959; AAA_ATPase_cent.
DR InterPro; IPR004605; RuvB.
DR Pfam; PF00004; AAA; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRfam; TIGR00635; ruvB; 1.
DR DNA repair; SOS response; ATP-binding; DNA recombination; Helicase;
KW Complete proteome.
FT NP BIND 45 52 ATP (POTENTIAL).
SQ SEQUENCE 307 AA; 35000 MW; 376E5138279E396C CRC64;

Query Match 11.8%; Score 100; DB 1; Length 307;
Best Local Similarity 26.5%; Pred. No. 0.07;
Matches 31; Conservative 27; Mismatches 39; Indels 20; Gaps 5;

Oy 11 NGEQVKGILYGPFGTGSFIIGAIANOLSKKVRSTIYLPFIETLKGF--KDGSPF 68
| : : | | | | : : | | : : | : : | : : | : : | : : | : : |
Db 34 NKTQLDHLILYGPFGVKTILARILNELKTK-----LQIQGHLQKPSDFL 81
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 69 KLIHRYEANIIMLDIGAEVTPWVRDEVIGPLHYRMVHELPTFPSSNFDYSLEL 125
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 82 NAIISLIKGVLEFIDEIHA--VAPNVM-ELMYPVMD---VFKIQVILGKDFNSKIYE 132
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: December 15, 2003, 15:14:18  
 Job time : 6.18845 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 15:11:40 ; Search time 18.9491 Seconds  
(without alignments)  
2235.384 Million cell updates/sec

Title: US-09-689-952-16

Perfect score: 851  
Sequence: 1 AADDICTAITNGEQVKGVL.....ERVKSLSTPYFLSGENFRNN 164

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

SPTREMBL\_23:\*  
1: sp archaea:\*  
2: sp bacteria:\*  
3: sp fungi:\*  
4: sp human:\*  
5: sp invertebrate:\*  
6: sp mammal:\*  
7: sp mhc:\*  
8: sp organelle:\*  
9: sp phage:\*  
10: sp plant:\*  
11: sp rodent:\*  
12: sp virus:\*  
13: sp vertebrate:\*  
14: sp unclassified:\*  
15: sp rvirus:\*  
16: sp bacteriaph:\*  
17: sp archaeop:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	851	100.0	306	16	Q99TH8
2	851	100.0	306	16	Q8NW67
3	782	91.9	306	16	Q8CNY2
4	423.5	49.8	307	16	Q92BF4
5	420.5	49.4	311	16	Q9K863
6	418.5	49.2	308	16	Q8CX80
7	418.5	49.2	307	16	Q8Y6X1
8	351	41.2	293	16	Q8P2J5
9	351	41.2	300	16	Q9A1D1
10	347	40.8	298	16	Q8CWP7
11	347	40.8	299	16	Q8CWP7
12	346	40.7	298	16	Q97PC7
13	332.5	39.1	300	16	Q8S3T8
14	330	38.8	293	16	Q9CHH9
15	330	38.8	293	16	Q9CHH9
16	205.5	24.1	74	2	Q05653

17	194	22.8	343	16	Q8EMKO	Q8EWKO mycoplasma
18	166	19.5	209	2	Q48991	Q48991 mycoplasma
19	149	17.5	297	2	Q98F03	Q98F03 mycoplasma
20	140.5	16.5	282	16	Q97H55	Q97H55 clostridium
21	140	16.5	316	16	Q9PQ14	Q9PQ14 ureaplasma
22	129.5	15.2	235	16	Q67056	Q67056 aquifex aeo
23	125.5	14.7	285	16	Q8E296	Q8E296 leptospira
24	122.5	14.4	259	9	Q9G022	Q9G022 bacterioph
25	118.5	13.9	259	9	Q9B0F8	Q9B0F8 staphylococ
26	116	13.6	261	16	Q932A3	Q932A3 staphylococ
27	113.5	13.3	195	17	Q8PTM8	Q8PTM8 methanosarc
28	112.5	13.2	257	17	Q9WBR8	Q9WBR8 staphylococ
29	112.5	13.2	257	17	Q8TRM8	Q8TRM8 methanosarc
30	112.5	13.2	257	17	Q8TH29	Q8TH29 methanosarc
31	111	13.0	350	16	Q921Y8	Q921Y8 rickettsia
32	110	12.9	262	2	Q93P71	Q93P71 microscilla
33	109	12.8	316	2	P95446	P95446 pseudomonas
34	108.5	12.7	257	17	Q8RKP6	Q8RKP6 thermomane
35	108.5	12.7	257	17	Q8RKP6	Q8RKP6 methanosarc
36	107	12.6	249	2	Q93SD5	Q93SD5 escherichia
37	106.5	12.5	261	2	Q8G101	Q8G101 pseudomonas
38	106.5	12.5	1038	10	Q9F856	Q9F856 arabidopsis
39	105.5	12.4	248	16	Q8X576	Q8X576 escherichia
40	105.5	12.4	796	17	Q8TZV0	Q8TZV0 pyrococcus
41	105	12.3	246	16	Q8XAD9	Q8XAD9 escherichia
42	105	12.3	375	17	Q97BR9	Q97BR9 thermoplasm
43	104.5	12.3	248	16	Q8X4T8	Q8X4T8 escherichia
44	104.5	12.3	257	17	Q8TLB4	Q8TLB4 methanosarc
45	104.5	12.3	287	16	Q926A4	Q926A4 listeria in

## ALIGNMENTS

RESULT 1	ID	Q99TH8	PRELIMINARY:	PRT:	306 AA.
AC	Q99TH8	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)				
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)				
DE	Phosomonal protein.				
GN	DNAI OR SAV1684 OR SA1507.				
OS	Staphylococcus aureus (strain Mu50 / ATCC 700699), and				
OS	Staphylococcus aureus (strain N315).				
OC	Bacteria; Firmicutes; Bacillales; Staphylococcus.				
OX	NCBI TaxID=158879; 158879;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S. aureus (strain Mu50), and S. aureus (strain N315);				
RX	MEDLINE=2111952; PubMed=11418146;				
RA	Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,				
RA	Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-Q., Ito T.,				
RA	Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,				
RA	Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,				
RA	Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,				
RA	Kanahisa M., Yamashita A., Oshtma K., Furuya K., Yoshino C., Shiba T.,				
RA	Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;				
RT	"Whole genome sequencing of methicillin-resistant Staphylococcus				
RT	aureus";				
RL	Lancet 357:1225-1240(2001).				
DR	EMBL; AP003363; BAB57846.1; -.				
DR	EMBL; AP003134; BAB42774.1; -.				
KW	Complete proteome.				
SC	SEQUENCE 306 AA; 35635 MW; 7F3440E89643505E CRC64;				
Query Match	100.0%; Score 851; DB 16; Length 306;				
Best Local Similarity	Pred. No. 5.6e-73;				
Matches 164; Conservative	0; Mismatches 0; Indels 0; Gaps 0;				
Qy	1 AADDICTAITNGEQVKGVLVGPFGTGSFTLGAIALNQLSKKRVSTIIYVPEFRTLKG 60				
Db	143 AADDICTAITNGEQVKGVLVGPFGTGSFTLGAIALNQLSKKRVSTIIYVPEFRTLKG 202				

Qy 61 GFKDGFPEKTLHRRREANITMLDDIGAEEVTPWVRDEVIGPLHYRMVHELPTFFSSNPD 120  
 Db 203 GFKDGFPEKTLHRRREANITMLDDIGAEEVTPWVRDEVIGPLHYRMVHELPTFFSSNPD 262

Qy 121 YSELEHHLAMTRDGEKTKAARIETRVKSLSTPYFLSGENFRNN 164  
 Db 263 YSELEHHLAMTRDGEKTKAARIETRVKSLSTPYFLSGENFRNN 306

## RESULT 2

Q8NM67 PRELIMINARY; PRT; 306 AA.  
 AC Q8NM67;  
 DT 01-OCT-2002 (TREMblrel. 22, Created)  
 DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)  
 DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)  
 DE Primosomal protein.  
 GN DNAI OR MW1627.  
 OS Staphylococcus aureus (strain MW2).  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OX NCBI\_TaxID=196620;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22040717; PubMed=12044378;  
 RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A., Nagai Y., Iwama N., Asano K., Naito T., Kuroda H., Cui L., Yamamoto K., Hiramatsu K.,  
 RT "Genome and virulence determinants of high virulence community-acquired MRSA."  
 RL Lancet 359:1819-1827 (2002).  
 DR EMBL, AP004827; BAB95492.1; -.  
 KM Complete proteome.  
 SQ SEQUENCE 306 AA; 35626 MW; COFCA3752D934B4 CRC64;

Query Match 100.0%; Score 851; DB 16; Length 306;  
 Best Local Similarity 100.0%; Pred. No. 5, 6e-73;  
 Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AADICTAINTNGEQVGLVYGPFGTGSFLIGAIANOLSKKVRSTIYLPFIRTLKG 60  
 Db 143 AADICTAINTNGEQVGLVYGPFGTGSFLIGAIANOLSKKVRSTIYLPFIRTLKG 202

Qy 61 GFKDGFPEKTLHRRREANITMLDDIGAEEVTPWVRDEVIGPLHYRMVHELPTFFSSNPD 120  
 Db 203 GFKDGFPEKTLHRRREANITMLDDIGAEEVTPWVRDEVIGPLHYRMVHELPTFFSSNPD 262

Qy 121 YSELEHHLAMTRDGEKTKAARIETRVKSLSTPYFLSGENFRNN 164  
 Db 263 YSELEHHLAMTRDGEKTKAARIETRVKSLSTPYFLSGENFRNN 306

## RESULT 3

Q8CNV2 PRELIMINARY; PRT; 306 AA.  
 AC Q8CNV2;  
 DT 01-MAR-2003 (TREMblrel. 23, Created)  
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
 DE Primosomal protein.  
 GN SEI358.  
 OS Staphylococcus epidermidis.  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OX NCBI\_TaxID=1282;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=ATCC 12228;  
 RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z., Chen Z., Wen Y.;  
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, AB016748; AA004957.1; -.  
 KM Complete proteome.  
 SQ SEQUENCE 306 AA; 35736 MW; EDD9F7B52DBFCA16 CRC64;

Query Match 91.9%; Score 782; DB 16; Length 306;  
 Best Local Similarity 89.6%; Pred. No. 2, 1e-66;  
 Matches 147; Conservative 12; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AADICTAINTNGEQVGLVYGPFGTGSFLIGAIANOLSKKVRSTIYLPFIRTLKG 60  
 Db 143 AADICTAINTNGEQVGLVYGPFGTGSFLIGAIANOLSKKVRSTIYLPFIRTLKG 202

Qy 61 GFKDGFPEKTLHRRREANITMLDDIGAEEVTPWVRDEVIGPLHYRMVHELPTFFSSNPD 120  
 Db 203 GFKDGFPEKTLHRRREANITMLDDIGAEEVTPWVRDEVIGPLHYRMVHELPTFFSSNPD 262

Qy 121 YSELEHHLAMTRDGEKTKAARIETRVKSLSTPYFLSGENFRNN 164  
 Db 263 YSELEHHLAMTRDGEKTKAARIETRVKSLSTPYFLSGENFRNN 306

## RESULT 4

Q92BF4 PRELIMINARY; PRT; 307 AA.  
 AC Q92BF4;  
 DT 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
 DE Primosome component (Helicase loader) DnaI.  
 GN DNAI OR LIN1595.  
 OS Listeria innocua.  
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
 OX NCBI\_TaxID=1642;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CLIP 11262 / Serovar 6a;  
 RX PubMed=11679669;  
 RA Glaeser P., Francau L., Buchrieser C., Ruehli C., Amend A., Baquero F., Berche P., Bloecher H., Brandt P., Chakraborty T., Charbit A., Chetoui F., Couve E., de Daruvar A., Dehoux P., Doman E., Dominguez-Bernal G., Duchaud E., Durant L., Duesenget O., Entian K.-D., Feihl W., Gomez-Lopez N., Hain T., Haut U., Jackson D., Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kutrappat G., Madueno E., Maltouman A., Mata Vicente J., Ng B., Nedjati H., Nordstrek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R., Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A., Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;  
 RT "Comparative genomics of Listeria species."  
 RL Science 294:849-852 (2001).  
 DR EMBL, AL596169; CAC96826.1; -.  
 DR Listeria; LIN01595;  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR SMART: SM00382; AAA; 1.  
 KM Helicase; Complete proteome.  
 SQ SEQUENCE 307 AA; 35988 MW; A8517663BA8F123F CRC64;

Query Match 49.8%; Score 423.5; DB 16; Length 307;  
 Best Local Similarity 52.3%; Pred. No. 3, 3e-32;  
 Matches 80; Conservative 32; Mismatches 40; Indels 1; Gaps 1;

Qy 11 NGEQVGLVYGPFGTGSFLIGAIANOLSKKVRSTIYLPFIRTLKGFKDGFPEK 70  
 Db 154 NGEQVGLVYGPFGTGSFLIGAIANOLSKKVRSTIYLPFIRTLKGFKDGFPEK 213

Qy 71 LHRVREANITMLDDIGAEEVTPWVRDEVIGPLHYRMVHELPTFFSSNPDYSELEHHLAM 130  
 Db 214 LHRVREANITMLDDIGAEEVTPWVRDEVIGPLHYRMVHELPTFFSSNPDYSELEHHLAM 273

Qy 131 TRDG-EKTKAARIETRVKSLSTPYFLSGENFR 162  
 Db 274 AONGTEKTLHRRREANITMLDDIGAEEVTPWVRDEVIGPLHYRMVHELPTFFSSNPD 306

## RESULT 5

Q9K863

```

ID 09K863 PRELIMINARY; PRT; 311 AA.
AC 09K863;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Primosome component (Helicase loader)
GN DNAI OR BH3144.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.
NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Maeni N.,
  Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
  Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
  halodurans and genomic sequence comparison with Bacillus subtilis."
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL: AP001517; BAB06863.1; -.
DR InterPro: IPR003593; AAA_ATPase.
DR SMART: SM00382; AAA; 1.
KW Helicase; Complete proteome.
SQ SEQUENCE 311 AA; 35822 MW; 6AFB3073431363BB CRC64;

Query Match 49.4%; Score 420.5; DB 16; Length 311.
Best Local Similarity 50.0%; Pred. No. 6,5e-32;
Matches 77; Conservative 33; Mismatches 43; Indels 1; Gaps 1;

QY 12 GGVKLVYGPFGTSGFLIGALIANOLKSKVRSSTIYLPFRITLKGFGKDSFEKKL 71
DB 156 GGVKLVYGPFGTSGFLIGALIANOLKSKVRSSTIYLPFRITLKGFGKDSFEKKL 215
QY 72 HVRREANILMDDIGAEVTPWVRDEVIGPLHYRMVHLPFPSSNPDYSELEHILAMT 131
DB 216 DFKNAQVLIFFDIDIGAEVTPWVRDEVIGPLHYRMVHLPFPSSNPDYSELEHILAMT 275
QY 132 -RDGEKTKAARIIEKVKSLSTPYPLSGENFRNN 164
DB 276 DKSGTELLAKRKYMERIRHYTVSVWVGQGNRRRH 309

RESULT 6
Q8CX80 PRELIMINARY; PRT; 308 AA.
AC 08CX80;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Primosome component (Helicase loader).
GN DNAI OR OB2156.
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacilliales; Oceanobacillus.
NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HTE831 / DSM 14371 / JCM 11309;
RX MEDLINE=22220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
  Ridge and its unexpected adaptive capabilities to extreme
  environments."
RL Nucleic Acids Res. 30:3927-3935(2002).
DR EMBL: AP004600; BAC14112.1; -.
KW Helicase; Complete proteome.
SQ SEQUENCE 308 AA; 35729 MW; 8D4C2F70E480A33 CRC64;

Query Match 49.4%; Score 420; DB 16; Length 308;
Best Local Similarity 53.4%; Pred. No. 7,2e-32;
Matches 79; Conservative 30; Mismatches 39; Indels 0; Gaps 0;

QY 16 KGLYVGFPGTSGFLIGALIANOLKSKVRSSTIYLPFRITLKGFGKDSFEKKLHRYA 75

```

```

DB 161 KGLYVGFPGTSGFLIGALIANOLKSKVRSSTIYLPFRITLKGFGKDSFEKKVDFPK 220
QY 76 EANILMDDIGAEVTPWVRDEVIGPLHYRMVHLPFPSSNPDYSELEHILAMTROE 135
DB 221 KADIIIMLDMDGMEKMSAMFRDEVLSVLYGRWMEGLPVRITSNYDLDLOELSTRNGV 280
QY 136 EKTAKARIIEKVKSLSTPYPLSGENFRNN 163
DB 281 EGVKGRILLERIKQVTTDVKLSGPNRRS 308

RESULT 7
Q8Y6X1 PRELIMINARY; PRT; 307 AA.
AC 08Y6X1;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Primosome component (Helicase loader) Dnai.
GN DNAI OR LM01560.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacilliales; Listeriaceae; Listeria.
NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BGD-e / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaeser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
  Baquero F., Frangeul P., Bloecher H., Brandt P., Chakraborty T.,
  Charbit A., Chetoui A., Couve E., de Darvar A., Deloux P.,
  Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussauget O.,
  Entian K.-D., Feibi H., Garcia-del Portillo F., Garrido P.,
  Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
  Jones L.-M., Kaerst U., Kretz J., Kuhn M., Kunet F., Kurapkat G.,
  Madueno E., Maitouram A., Mata Vicente J., Ng B., Nedjari H.,
  Nordiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
  Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
  Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomes of Listeria species."
RL Science 294:849-852(2001).
DR EMBL: AL591979; CAC99638.1; -.
DR Listlist; LM001560; -.
DR InterPro: IPR003593; AAA_ATPase.
DR SMART: SM00382; AAA; 1.
KW Helicase; Complete proteome.
SQ SEQUENCE 307 AA; 35970 MW; CF2FD9350CEB0F48 CRC64;

Query Match 49.2%; Score 418.5; DB 16; Length 307;
Best Local Similarity 51.6%; Pred. No. 9,9e-32;
Matches 79; Conservative 33; Mismatches 40; Indels 1; Gaps 1;

QY 11 NGBOVKLVYGPFGTSGFLIGALIANOLKSKVRSSTIYLPFRITLKGFGKDSFEKK 70
DB 154 SGEVYVGLTFHSGFGKSYLLGALAKELALGISTTVLYLPFRMREKVSISDNTVGRK 213
QY 71 LHRVREANILMDDIGAEVTPWVRDEVIGPLHYRMVHLPFPSSNPDYSELEHILAM 130
DB 214 IOPAKETEVLMDDIGAEVTPWVRDEVIGPLHYRMVHLPFPSSNPDYSELEHILAM 273
QY 131 TRDG-EKTKARIIEKVKSLSTPYPLSGENFR 162
DB 274 AONGTEKAKARIIEKVKSLSTPYPLSGENFR 306

RESULT 8
Q8P2J5 PRELIMINARY; PRT; 293 AA.
AC 08P2J5;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Putative primosome component (helicase loader).

```

GN SPY18.0332.  
 OS Streptococcus pyogenes (serotype M18).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 RN NCBI\_TaxID=186103;  
 [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=MGAS8232 / Serotype M18;  
 RX MEDLINE=21927593; PubMed=11917108;  
 RA Smoot J.C., Bardian K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,  
 RA Sylva G.L., Studevant D.E., Ricklets S.M., Porcella S.F.,  
 RA Perkins L.D., Bares S.B., Campbell D.S., Smith T.M., Zhang Q.,  
 RA Kapur V., Daly J.A., Veasy L.G., Mueser J.M.;  
 RT "Genome sequence and comparative microarray analysis of serotype M18  
 RT group A Streptococcus strains associated with acute rheumatic fever  
 RT outbreaks";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).  
 KM EMBL; AB009978; AL97087.1; -  
 SQ SEQUENCE 293 AA; 33357 MW; 087AD94BDB59986 CRC64;  
 Query Match. 41.2%; Score 351; DB 16; Length 293;  
 Best Local Similarity 43.1%; Pred. No. 2.6e-25;  
 Matches 69; Conservative 36; Mismatches 53; Indels 2; Gaps 2;  
 QY 4 DICTAINGOVKGLYGPFGTGSFLIGAINOLSKK-VSTIYLPFRITLKGCF 62  
 DB 135 DPEYQVSAEQ-KGLYIGDMGKSYLLAAHHESEKGVSTLLHPSPFAIDVKNAI 193  
 QY 63 KDSFEKKLRHREANILMLDDIGAEVTPWRDEVIQPLHYRMVHELPTFSSNPDYS 122  
 DB 194 SNGSVKEIDAVKVNVLILDDIGAEQATSWNREVLQVILQYRMLELPTFTSNVSPA 253  
 QY 123 ELEHHLAMTRDGEKTKAARIIRVKSISTPYPLSGENFR 162  
 DB 254 DLERKVAITIKGSDETWQAKRWERVRYTLAREFHLEGANRR 293  
 RESULT 9  
 Q9A1D1 PRELIMINARY; PRT; 300 AA.  
 ID Q9A1D1  
 AC Q9A1D1  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
 DE Putative pilin component (Helicase loader).  
 GN DNA1 OR SPY0340 OR SPY03.0248.  
 OS Streptococcus pyogenes, and  
 OS Streptococcus pyogenes (serotype M3).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OC NCBI\_TaxID=1314, 198466;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP SPECIES=pyogenes; STRAIN=SF370 / ATCC 700294 / Serotype M1;  
 RX MEDLINE=21192684; PubMed=12196296;  
 RA Perrecci J.J., McShan W.M., Ajdic D.J., Savic D.J., Saye G., Lyon K.,  
 RA Primaud C., Szate S., Suvorov A.N., Kenon S., Lai H.S., Lin S.P.,  
 RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,  
 RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;  
 RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RP SPECIES=pyogenes; STRAIN=MGAS315 / Serotype M3;  
 RX MEDLINE=22133808; PubMed=12122206;  
 RA Bares S.B., Sylva G.L., Bardian K.D., Lei B., Hoff J.S.,  
 RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Perkins L.D.,  
 RA Campbell D.S., Smith T.M., McCormick U.K., Leung D.Y.M.,  
 RA Schlievert F.M., Mueser J.M.;  
 RT "Genome sequence of a serotype M3 strain of group A Streptococcus:  
 RT phage-encoded toxins, the high-virulence phenotype, and clone  
 RT emergence.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).  
 DR EMBL; AB006498; AK33392.1; -  
 DR EMBL; AB014141; AM78855.1; -  
 KM Hypothetical protein; Helicase; Complete proteome.  
 SQ SEQUENCE 300 AA; 34146 MW; 489B27A789338C5A CRC64;  
 Query Match. 41.2%; Score 351; DB 16; Length 300;  
 Best Local Similarity 43.1%; Pred. No. 2.6e-25;  
 Matches 69; Conservative 36; Mismatches 53; Indels 2; Gaps 2;  
 QY 4 DICTAINGOVKGLYGPFGTGSFLIGAINOLSKK-VSTIYLPFRITLKGCF 62  
 DB 142 DPEYQVSAEQ-KGLYIGDMGKSYLLAAHHESEKGVSTLLHPSPFAIDVKNAI 200  
 QY 63 KDSFEKKLRHREANILMLDDIGAEVTPWRDEVIQPLHYRMVHELPTFSSNPDYS 122  
 DB 201 SNGSVKEIDAVKVNVLILDDIGAEQATSWNREVLQVILQYRMLELPTFTSNVSPA 260  
 QY 123 ELEHHLAMTRDGEKTKAARIIRVKSISTPYPLSGENFR 162  
 DB 261 DLERKVAITIKGSDETWQAKRWERVRYTLAREFHLEGANRR 300  
 RESULT 10  
 Q8CWP7 PRELIMINARY; PRT; 298 AA.  
 ID Q8CWP7  
 AC Q8CWP7  
 DT 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Pilin component (Helicase loader).  
 GN DNA1 OR SPY155.  
 OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OC NCBI\_TaxID=171101;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP MEDLINE=21429245; PubMed=11544234;  
 RA Hoskins J., Aldorn W.B., Jr., Arnold J., Blaszcak L.C., Burgett S.,  
 RA DeHoff B.S., Battem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,  
 RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,  
 RA LeBlanc D.J., Lee L.N., Lefkowitz S.J., Lu J., Matsushima P.,  
 RA McAben S.M., McManney M., Mcleaster K., Mundy C.W., Nicas T.I.,  
 RA Norris P.H., O'Gara M., Peery R.B., Robertson G.T., Rokey P.,  
 RA Sun P.-M., Winkler M.B., Yang Y., Young-Bellido M., Zhao G.,  
 RA Zook C.A., Baltz R.H., Jaskunas S.R., Kostek P.R., Jr., Skatrud P.L.,  
 RA "Genome of the bacterium Streptococcus pneumoniae strain R6";  
 RT J. Bacteriol. 183:5709-5717(2001).  
 RL EMBL; AB008523; AAL00359.1; -  
 KM Helicase; Complete proteome.  
 SQ SEQUENCE 298 AA; 34574 MW; 3EA8DB8BD1DEB22 CRC64;  
 Query Match. 40.8%; Score 347; DB 16; Length 298;  
 Best Local Similarity 44.0%; Pred. No. 6.3e-25;  
 Matches 66; Conservative 36; Mismatches 46; Indels 2; Gaps 2;  
 QY 15 VGLYLYGPFGTGSFLIGAINOLSKK-VSTIYLPFRITLKGCFKDSFEKKLR 73  
 DB 149 LKGLYIGDFGVGKSPVAAALAHDLSEKRGVSTLLHPSPFYDVAIDNAGVTLVD8 208  
 QY 74 VREANILMLDDIGAEVTPWRDEVIQPLHYRMVHELPTFSSNPDYSLEHHLAMTRD 133  
 DB 209 IKLSEVLLILDDIGAEQATSWNREVLQVILQYRMLELPTFTSNVSPA 268  
 QY 134 GEKTKAARIIRVKSISTPYPLSGENFR 162  
 DB 269 GNDETWAKRWERVRYTLAREFHLEGANRR 298  
 RESULT 11  
 Q8CWP7

ID 08CWM7 PRELIMINARY; PRT; 299 AA.  
 AC 08CWM7;  
 DT 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Putative DNA replication protein, primosome component (Helicase loader)  
 DE DNAI OR SMC1.921.  
 GN Streptococcus mutans.  
 OS Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=UA159 / ATCC 700610 / Serotype C;  
 RX MEDLINE=22295063; PubMed=12397186;  
 RA Ajdic D., Mcshan W.M., McLaughlin R.E., Savic G., Chang J.,  
 RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,  
 RA Li S., Zhu H., Najjar P., Lai H., White J., Roe B.A., Perletti J.J.;  
 RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental pathogen."  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).  
 DR EMBL; AE015016; AAN59532.1;  
 KW Helicase; Complete proteome.  
 SQ SEQUENCE 299 AA; 34410 MW; 98CAAB56F5524BD1 CRC64;  
 Query Match 40.8%; Score 347; DB 16; Length 299;  
 Best Local Similarity 42.5%; Pred. No. 6.3e-25;  
 Matches 68; Conservative 37; Mismatches 53; Indels 2; Gaps 2;  
 QY 4 DICTAITNEOVKGLVYGPFGKSPFLIGAIANQLSKK-VRSSTIYLPFRITLKGFGFEKKLHRY 62  
 DB 141 DVEQYPPNEBQ-KGLVYGDWKGKSYLMAAMHLSBQGAATFLHFPSTFIDVKNAI 199  
 QY 63 KDSEFEKLVREANITLMDDIGAEVTPWVDEVIQPLHYRMVHLPFFSSNPDYS 122  
 DB 200 NGTVKKEIDAVKATLILDDIGAGOSTSWINDEVLYQVLYQRMLELPFTFTSYSPK 259  
 QY 123 ELHHILMTDGEKTKAARIIRVKSISTPYFLSGENR 162  
 DB 260 DLEAKLANIKGSDETWQAKRMERIRYLAETLEGENNR 299  
 RESULT 12  
 Q97PC7 PRELIMINARY; PRT; 298 AA.  
 AC Q97PC7;  
 DT 01-OCT-2001 (TREMBLrel. 18, Created)  
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Primosomal protein DnaI.  
 GN SPI171.  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1313;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TIGR4;  
 RX MEDLINE=21357209; PubMed=11463916;  
 RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
 RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,  
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Raddue D.,  
 RA Holtzapfel E., Khouli H., Wolf A.M., Ueberback T.R., Hansen C.L.,  
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickenson T., Hickey E.K.,  
 RA Holt I.B., Loftus B.J., Yang F., Smith H.O., Venter J.C., Venter C.M.;  
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Frazer C.M.;  
 RT "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae."  
 RT Science 293:498-506(2001).  
 DR EMBL; AE007464; AAK75789.1;  
 TIGR, SPI171;

KW Complete proteome.  
 SQ SEQUENCE 298 AA; 34259 MW; CA5EBCAFA9C30059 CRC64;  
 Query Match 40.7%; Score 346; DB 16; Length 298;  
 Best Local Similarity 44.3%; Pred. No. 7.8e-25;  
 Matches 66; Conservative 35; Mismatches 46; Indels 2; Gaps 2;  
 QY 16 KGLVYGPFGKSPFLIGAIANQLSKK-VRSSTIYLPFRITLKGFGFGFEKKLHRY 74  
 DB 150 KGLVYGDWKGKSPVVAALAHDLSEKGVSTLHFPSTFIDVKNAISDGVKTLVDEI 209  
 QY 75 REANITLMDDIGAEVTPWVDEVIQPLHYRMVHLPFFSSNPDYSLEHHILMTDGG 134  
 DB 210 KLSVLLIIDDIGAEVSTWVNDLQVLYQVMOENLPFTFTSNFDELEHGPAPKVKHG 269  
 QY 135 EERT-KAARIIRVKSISTPYFLSGENR 162  
 DB 270 NDETWEARVMERIRYLAETLEGENNR 298  
 RESULT 13  
 Q9E3T8 PRELIMINARY; PRT; 300 AA.  
 AC Q9E3T8;  
 DT 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 GN DNAI OR GBS1668.  
 OS Streptococcus agalactiae (serotype III).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=216495;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MEM316 / Serotype III;  
 RX MEDLINE=22242508; PubMed=12354421;  
 RA Glaeser P., Ruenick C., Buchrieser C., Chevalier F., Frangoul L.,  
 RA Maeder T., Zouine M., Couve E., Lallouli H., Poyart C., Tieu-Cuot P.,  
 RA Kunat F.;  
 RT "Genome sequence of Streptococcus agalactiae, a pathogen causing invasive neonatal disease."  
 RT Mol. Microbiol. 45:1499-1513 (2002).  
 DR EMBL; AL766852; CAD47327.1;  
 DR Sagaliet; GBS1668;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 300 AA; 34679 MW; 9C0B4898661661E9 CRC64;  
 Query Match 39.1%; Score 332.5; DB 16; Length 300;  
 Best Local Similarity 41.9%; Pred. No. 1.5e-23;  
 Matches 62; Conservative 36; Mismatches 49; Indels 1; Gaps 1;  
 QY 16 KGLVYGPFGKSPFLIGAIANQLSKK-VRSSTIYLPFRITLKGFGFGFEKKLHRY 74  
 DB 153 KGLVYGDWKGKSPVVAALAHDLSEKGVSTLHFPSTFIDVKNAISDGVKTLVDEI 212  
 QY 75 REANITLMDDIGAEVTPWVDEVIQPLHYRMVHLPFFSSNPDYSLEHHILMTDGG 134  
 DB 213 KSVPLIIDDIGAEVSTWVNDLQVLYQVMOENLPFTFTSNFDELEHGPAPKVKHG 272  
 QY 135 EERT-KAARIIRVKSISTPYFLSGENR 162  
 DB 273 DETWQAKRMERIRYLAETLEGENNR 300  
 RESULT 14  
 Q9DY72 PRELIMINARY; PRT; 300 AA.  
 AC Q9DY72;  
 DT 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Primosomal protein DnaI.

GN DNAI OR SAG1621.  
 OS Streptococcus agalactiae (serotype V).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 NCBI\_TaxID=216466;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=2603 V/R / Serotype V;  
 RX MEDLINE=2222298; PubMed=12200547;  
 RA Tettein H., Maignan V., Cieslewicz M.J., Eisen J.A., Peterson S.,  
 RA Wessels M.R., Paulsen I.T., Nelson K.E., Margalit I., Read T.D.,  
 RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,  
 RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,  
 RA Radu D., Fedorova N.B., Scanlan D., Khouli H., Mulligan S.,  
 RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarelli M., Mora M.,  
 RA Jacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Malone D.,  
 RA Rinaldo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,  
 RA Fraser C.M.;  
 RT "Complete genome sequence and comparative genomic analysis of an  
 RT emerging human pathogen, serotype V Streptococcus agalactiae";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).  
 DR EMBL: AE014265; AAN00485.1; -.  
 DR TIGR: SAG1621; -.  
 KW Complete proteome.  
 SQ SEQUENCE 300 AA; 34663 MW; 9BED36F86B36B3E9 CRC64;

Query Match 39.1%; Score 332.5; DB 16; Length 300;  
 Best Local Similarity 41.9%; Pred. No. 1.5e-23;  
 Matches 62; Conservative 36; Mismatches 49; Indels 1; Gaps 1;

QY 16 KGLYVGPFGTGSFLIGAIANOLKSKK-VRSITIIYLPPIRTLKGFGKDSFEKGLHRY 74  
 Db 153 KGLYVGDWGVGSKYLMAMARELSERKGVSTILLHPSPALDVKNALSSGTVDIDAV 212  
 QY 75 REANITLMDIGAEVTPWVRDEVIGPLHYRWVHELPTFSSNPDYSELEHLAMTRDG 134  
 Db 213 KSVPIILLDDIGABQATSWVRDELQVILQHRMLELPTFTSNYSFNDLERKMANIKGS 272  
 QY 135 EEKTKARITERYKSLSTPYFLSGENFR 162  
 Db 273 DETWQAKRVERRYLAIEFHLGPNRR 300

## RESULT 15

Q9CHH9 PRELIMINARY; PRT; 293 AA.  
 ID Q9CHH9  
 AC Q9CHH9;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Primosomal protein DnaI.  
 GN DNAI OR IL0752.  
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.  
 NCBI\_TaxID=1360;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IL1403;  
 RX MEDLINE=21235186; PubMed=11337471;  
 RA Bolotin A., Wincker P., Manger S., Jallion O., Malarne K.,  
 RA Weisenbach J., Ehrlich S.D., Sorokin A.;  
 RT "The complete genome sequence of the lactic acid bacterium Lactococcus  
 RT lactis ssp. lactis IL1403."  
 RL Genome Res. 11:731-753(2001).  
 DR EMBL: AE006308; AAK04850.1; -.  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR SMART: SM00382; AAA; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 293 AA; 33904 MW; D775C530A69445B3 CRC64;

Query Match 38.8%; Score 330; DB 16; Length 293;  
 Best Local Similarity 43.4%; Pred. No. 2.6e-23;  
 Matches 69; Conservative 30; Mismatches 50; Indels 10; Gaps 2;

QY 9 ITNGEQVKGILYVGPFGTGSFLIGAIANOLKSKKVRSTIILYLPPIRTLKGFGKDSFE 68  
 Db 142 ITNPPKKKGLYVGDWGVGSKSFLMAMARELSERKGVSTILLHPSPALDVKNALSSGTVDIDAV 194  
 QY 69 KK---LHRYREANITLMDIGAEVTPWVRDEVIGPLHYRWVHELPTFSSNPDYSELE 125  
 Db 195 NAKVWVNEIKASQVILVDDIGAEQNNAMVRDSILQVILQHRMQENLPTFTSNLRLMELE 254  
 QY 126 HHLAMTRDGEKTKARITERYKSLSTPYFLSGENFRN 164  
 Db 255 QHLAETKRADEIWPARKVMERVRYLAIEEKLSTGTRRD 293

Search completed: December 15, 2003, 15:15:57  
 Job time : 19.9491 secs

XX

PT preferably bacterial, diseases such as those caused by Staphylococcus  
 PT aureus -  
 XX  
 PS Disclosure; Fig 15; 107bp; English.

CC This sequence shows a DnaI polypeptide derived from S. aureus. S.  
 CC aureus DnaI interacts with a growth-inhibitory bacteriophage 77 ORF  
 CC 104 gene product, to form the basis of a screening assay. DnaI  
 CC polypeptides and polynucleotides are useful for treating microbial,  
 CC preferably bacterial, especially Staphylococcal, infections. DnaI  
 CC polypeptides and polynucleotides are useful for biological, diagnostic,  
 CC prophylactic, clinical and therapeutic use, and as components in  
 CC databases useful for search analyses as well as in sequence analysis  
 CC algorithms.

XX  
 SQ Sequence 250 AA;

Query Match 100.0%; Score 1328; DB 22; Length 250;  
 Best Local Similarity 100.0%; Pred. No. 3.6e-129;  
 Matches 250; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YKQOQHVDGKRPADCPNFKGHPVPELYVNNRIKIRYLQCPCKIKYDEERFEAEITSH 60  
 DB 1 YKQOQHVDGKRPADCPNFKGHPVPELYVNNRIKIRYLQCPCKIKYDEERFEAEITSH 60  
 QY 61 HMQDRLNAKLDKIYNNHRLDVMAADICTAINTGEOVKLYLGPFGTGSFTLGA 120  
 DB 61 HMQDRLNAKLDKIYNNHRLDVMAADICTAINTGEOVKLYLGPFGTGSFTLGA 120  
 QY 121 IANOLSKKVRSTIYLPFIRITLKGFGKDSPEKCLHVRANILMLDDIGAEVTPWV 180  
 DB 121 IANOLSKKVRSTIYLPFIRITLKGFGKDSPEKCLHVRANILMLDDIGAEVTPWV 180  
 QY 181 RDEVIGPLHYRWVHELPTFFSSNFDYSELEHLLAMTRDGEKTKARIIERYKSLSTPY 240  
 DB 181 RDEVIGPLHYRWVHELPTFFSSNFDYSELEHLLAMTRDGEKTKARIIERYKSLSTPY 240  
 QY 241 FLSGENFRNN 250  
 DB 241 FLSGENFRNN 250

# RESULT 2

AAB47317  
 ID AAB47317 standard; Protein; 313 AA.

AC AAB47317;

DT 29-AUG-2001 (first entry)

XX S. aureus DnaI.

KW DnaI, S. aureus; inhibitor; bacteriophage 77; ORF 104; phage 77ORF104;  
 screening assay.

OS Staphylococcus aureus.

XX MO200146383-A2.

XX 28-JUN-2001.

PF 21-DEC-2000; 2000MO-US35180.

XX 22-DEC-1999; 99US-0470512.

PR 12-OCT-2000; 2000US-0689952.

XX (PHAG-) PHAGETECH INC.

PA (WILL/) WILLIAMS K M.

XX Pelletier J, Gros P, Dubow M;

XX MPI; 2001-418052/44.

DR N-PSDB; AAC86105.

XX Novel DnaI polypeptides useful for treating and diagnosing microbial,  
 PT preferably bacterial, diseases such as those caused by Staphylococcus  
 PT aureus -  
 XX  
 PS Disclosure; Fig 1; 107bp; English.

CC This sequence shows DnaI derived from S. aureus. S. aureus DnaI  
 CC interacts with a growth-inhibitory bacteriophage 77 ORF 104  
 CC gene product, to form the basis of a screening assay. DnaI  
 CC polypeptides and polynucleotides are useful for treating microbial,  
 CC preferably bacterial, especially Staphylococcal, infections. DnaI  
 CC polypeptides and polynucleotides are useful for biological, diagnostic,  
 CC prophylactic, clinical and therapeutic use, and as components in  
 CC databases useful for search analyses as well as in sequence analysis  
 CC algorithms.

XX  
 SQ Sequence 313 AA;

Query Match 100.0%; Score 1328; DB 22; Length 313;  
 Best Local Similarity 100.0%; Pred. No. 4.9e-129;  
 Matches 250; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YKQOQHVDGKRPADCPNFKGHPVPELYVNNRIKIRYLQCPCKIKYDEERFEAEITSH 60  
 DB 64 YKQOQHVDGKRPADCPNFKGHPVPELYVNNRIKIRYLQCPCKIKYDEERFEAEITSH 123  
 QY 61 HMQDRLNAKLDKIYNNHRLDVMAADICTAINTGEOVKLYLGPFGTGSFTLGA 120  
 DB 124 HMQDRLNAKLDKIYNNHRLDVMAADICTAINTGEOVKLYLGPFGTGSFTLGA 183  
 QY 121 IANOLSKKVRSTIYLPFIRITLKGFGKDSPEKCLHVRANILMLDDIGAEVTPWV 180  
 DB 184 IANOLSKKVRSTIYLPFIRITLKGFGKDSPEKCLHVRANILMLDDIGAEVTPWV 243  
 QY 181 RDEVIGPLHYRWVHELPTFFSSNFDYSELEHLLAMTRDGEKTKARIIERYKSLSTPY 240  
 DB 244 RDEVIGPLHYRWVHELPTFFSSNFDYSELEHLLAMTRDGEKTKARIIERYKSLSTPY 303  
 QY 241 FLSGENFRNN 250  
 DB 304 FLSGENFRNN 313

# RESULT 3

AAG82049  
 ID AAG82049 standard; Protein; 299 AA.

AC AAG82049;

DT 03-SEP-2001 (first entry)

XX S. epidermidis open reading frame protein sequence SEQ ID NO:1192.

KW Staphylococcus epidermidis SRI strain; infection; diagnosis;  
 vaccination; endocarditis.

XX Staphylococcus epidermidis.

XX MO200134809-A2.

XX 17-MAY-2001.

PF 09-NOV-2000; 2000MO-US30782.

XX 09-NOV-1999; 99US-0164258.

XX (GLAXO ) GLAXO GROUP LTD.

PA kimmerly WJ;

XX MPI; 2001-316495/33.

DR N-PSDB; AAG52899.

XX Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*,  
 PT useful for vaccinating against infections, e.g. endocarditis -  
 XX  
 PS Claim 18; Page 342; 2189pp; English.

XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
 CC (II), given in AG81454 to AG83120, from *Staphylococcus epidermidis*.  
 CC (I) and (II) can have antibacterial activity and therefore can be used  
 CC in vaccination. The nucleic acids (I) may be used to produce the  
 CC S. epidermidis polypeptides (II) via the production of vectors  
 CC containing them which are used to produce host cells which express the  
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
 CC used to vaccinate subjects and to raise antibodies against the bacteria.  
 CC The polypeptides may also be used to assay for other inhibitors of their  
 CC activity and therefore identify compounds that may be used for the  
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
 CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
 CC polynucleotide sequences from the present invention. AAH55091 to  
 CC AAH55098 represent oligonucleotide sequences and primers which are used  
 CC in the exemplification of the present invention.  
 CC N.B. The present invention specifically claims all the polynucleotide  
 CC sequences given in the sequence listing of the present specification,  
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
 CC no sequences are present for SEQ ID NO:4455 to 4464.

XX Sequence 299 AA;

Query Match 89.5%; Score 1189; DB 22; Length 299;  
 Best Local Similarity 87.2%; Pred. No. 1.2e-114;  
 Matches 218; Conservative 23; Mismatches 9; Indels 0; Gaps 0;

QY 1 YDQQRKHYDGHKFPADCPNFKVGHVPELYVDNNKIRIYLOCPCKIKYDEERREABLTSH 60  
 DB 50 YDQQRKHYDGHKFPADCPNFKVGHVPELYVDNNKIRIYLOCPCKIKYDEERREABLTSH 109  
 QY 61 HMQRDTLNAKLDIYNNRRDLVMAAADICTALTNGBOVGLYVGFPGKSFILGA 120  
 DB 110 HMQRDTLNAKLDIYNNRRDLVMAAADICTALTNDKRVKGLYVGFPGKSFILGA 169  
 QY 121 IANOLSKKVRSTIYLPFIRITLKGFGDSFEKKLQVRREANILMDIDGAEVTPMV 180  
 DB 170 IANOLSKKVRSTIYLPFIRITLKGFGDSFEKKLQVRREANILMDIDGAEVTPMV 229  
 QY 181 RDEVIGPLIHYRMVHELPTFFSSNFDYSELEHLAMTRDGEKTKAARIIEVKSLSTPY 240  
 DB 230 RDEVIGPLIHYRMVHELPTFFSSNFDYSELEHLAMTRDGEKTKAARIIEVKSLSTPY 289  
 QY 241 FLGSENFRRN 250  
 DB 290 YLTGKNFRRN 299

RESULT 4

AA82337  
 ID AA82337 standard; Protein; 299 AA.

XX AA82337;

DT 03-SEP-2001 (first entry)

DE S. epidermidis open reading frame protein sequence SEQ ID NO:1768.

KW *Staphylococcus epidermidis* SRI strain; infection; diagnosis;  
 KM vaccination; endocarditis.

OS *Staphylococcus epidermidis*.

XX WO200134809-A2.

PD 17-MAY-2001.

XX

PF 09-NOV-2000; 2000WO-US30782.

XX 09-NOV-1999; 99US-0164258.

PA (GLAX) GLAXO GROUP LTD.

PI Kimmery NJ;

XX WPI, 2001-316495/33.

DR N-PSDB; AAH53187.

PT Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*,  
 useful for vaccinating against infections, e.g. endocarditis -

PS Claim 18; Page 488; 2189pp; English.

XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
 CC (II), given in AG81454 to AG83120, from *Staphylococcus epidermidis*.  
 CC (I) and (II) can have antibacterial activity and therefore can be used  
 CC in vaccination. The nucleic acids (I) may be used to produce the  
 CC S. epidermidis polypeptides (II) via the production of vectors  
 CC containing them which are used to produce host cells which express the  
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
 CC used to vaccinate subjects and to raise antibodies against the bacteria.  
 CC The polypeptides may also be used to assay for other inhibitors of their  
 CC activity and therefore identify compounds that may be used for the  
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
 CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
 CC polynucleotide sequences from the present invention. AAH55091 to  
 CC AAH55098 represent oligonucleotide sequences and primers which are used  
 CC in the exemplification of the present invention.  
 CC N.B. The present invention specifically claims all the polynucleotide  
 CC sequences given in the sequence listing of the present specification,  
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
 CC no sequences are present for SEQ ID NO:4455 to 4464.

XX Sequence 299 AA;

Query Match 89.5%; Score 1189; DB 22; Length 299;  
 Best Local Similarity 87.2%; Pred. No. 1.2e-114;  
 Matches 218; Conservative 23; Mismatches 9; Indels 0; Gaps 0;

QY 1 YDQQRKHYDGHKFPADCPNFKVGHVPELYVDNNKIRIYLOCPCKIKYDEERREABLTSH 60  
 DB 50 YDQQRKHYDGHKFPADCPNFKVGHVPELYVDNNKIRIYLOCPCKIKYDEERREABLTSH 109  
 QY 61 HMQRDTLNAKLDIYNNRRDLVMAAADICTALTNGBOVGLYVGFPGKSFILGA 120  
 DB 110 HMQRDTLNAKLDIYNNRRDLVMAAADICTALTNDKRVKGLYVGFPGKSFILGA 169  
 QY 121 IANOLSKKVRSTIYLPFIRITLKGFGDSFEKKLQVRREANILMDIDGAEVTPMV 180  
 DB 170 IANOLSKKVRSTIYLPFIRITLKGFGDSFEKKLQVRREANILMDIDGAEVTPMV 229  
 QY 181 RDEVIGPLIHYRMVHELPTFFSSNFDYSELEHLAMTRDGEKTKAARIIEVKSLSTPY 240  
 DB 230 RDEVIGPLIHYRMVHELPTFFSSNFDYSELEHLAMTRDGEKTKAARIIEVKSLSTPY 289  
 QY 241 FLGSENFRRN 250  
 DB 290 YLTGKNFRRN 299

RESULT 5

ABP38947  
 ID ABP38947 standard; Protein; 307 AA.

XX ABP38947;

DT 24-JUL-2002 (first entry)

DE *Staphylococcus epidermidis* ORF amino acid sequence SEQ ID NO:3792.

XX

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
 KM antibacterial; gene therapy.  
 XX  
 OS Staphylococcus epidermidis.  
 XX  
 PN US6380370-B1.  
 XX  
 PD 30-APR-2002.  
 XX  
 PF 13-AUG-1998; 98US-0134001.  
 XX  
 PR 14-AUG-1997; 97US-055779P.  
 PR 08-NOV-1997; 97US-064964P.  
 XX  
 PA (GENO-) GENOME THERAPEUTICS CORP.  
 XX  
 PI Doucette-Stamm LA, Bush D;  
 XX  
 PI WPI; 2002-381255/41.  
 DR N-PSDB; ABN91432.  
 XX  
 DR Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
 XX polypeptide, useful for diagnosing and treating bacterial infections -  
 PT disclosure; SEQ ID 3792; 267pp; English.  
 XX  
 PS ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
 CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
 CC antibacterial activity and can be used in gene therapy. The sequences  
 CC can also be used in the diagnosis and treatment of bacterial infections,  
 CC particularly S. epidermidis infections. The sequences can be used to  
 CC screen for compounds able to interfere with the S. epidermidis life  
 CC cycle or inhibit S. epidermidis infection.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC USPTO web site.  
 XX  
 XX  
 SQ Sequence 307 AA;  
 Query Match 89.5%; Score 1189; DB 23; Length 307;  
 Best Local Similarity 87.2%; Pred. No. 1.2e-114;  
 Matches 218; Conservative 23; Mismatches 9; Indels 0; Gaps 0;  
 QY 1 YKQOQHYDGHKADCFNFKYGVPELYVNNRNIKIRYLOCPCKIKYDERFPAELITSH 60  
 DB 58 YKQOQHYDGHKADCFNFKYGVPELYVNNRNIKIRYLOCPCKIKYDERFPAELITSH 117  
 QY 61 HMQRDITLNAKLDKIYNNHRRDLVAMADICTAITNGEQVKGILYGPFGTGSFIIGA 120  
 DB 118 HMQRDITLNAKLDKIYNNHRRDLVAMADICTAITNGEQVKGILYGPFGTGSFIIGA 177  
 QY 121 IANQLSKKVSTIILYPERIRTLKGGFKDGFSEKILRVREANIIMLDDIGAEVTPWV 180  
 DB 178 IANQLSKKVSTIILYPERIRTLKGGFKDGFSEKILRVREANIIMLDDIGAEVTPWV 237  
 QY 181 RDEVIGPLHYRMVHELPTFSSNFDYSELHHLAMTRDGEKTKAARIIRVKSISTPY 240  
 DB 238 RDEVIGPLHYRMVHELPTFSSNFDYSELHHLAMTRDGEKTKAARIIRVKSISTPY 297  
 QY 241 FLGSENFRRN 250  
 DB 298 YLTGKNFRNN 307

XX Amino acids 150-313 of S. aureus DnaI.  
 DE DnaI; S. aureus; inhibitor; bacteriophage 77; ORF 104; phage 77ORF104;  
 XX screening assay.  
 KM  
 XX Staphylococcus aureus.  
 OS  
 XX WO200146383-A2.  
 PN  
 XX 28-JUN-2001.  
 PD  
 XX 21-DEC-2000; 2000WO-US35180.  
 PF  
 XX 22-DEC-1999; 99US-0470512.  
 PR 12-OCT-2000; 2000US-0689952.  
 XX  
 XX (PHAG-) PHAGETECH INC.  
 PA (WILL/) WILLIAMS K M.  
 XX  
 PI Pelletier J, Gros P, Dubow M;  
 XX  
 PI WPI; 2001-418052/44.  
 DR N-PSDB; AAC86104.  
 XX  
 DR Novel DnaI polypeptides useful for treating and diagnosing microbial,  
 PT preferably bacterial, diseases such as those caused by Staphylococcus  
 PT aureus -  
 XX  
 PS Claim 41; Fig 15; 107pp; English.  
 XX  
 XX This sequence shows a DnaI polypeptide derived from S. aureus. S.  
 CC aureus DnaI interacts with a growth-inhibitory bacteriophage 77 ORF  
 CC 104 gene product, to form the basis of a screening assay. DnaI  
 CC polypeptides and polynucleotides are useful for treating microbial,  
 CC preferably bacterial, especially Staphylococcus, infections. DnaI  
 CC polypeptides and polynucleotides are useful for biological, diagnostic,  
 CC prophylactic, clinical and therapeutic use, and as components in  
 CC databases useful for search analyses as well as in sequence analysis  
 CC algorithms.  
 XX  
 SQ Sequence 164 AA;  
 Query Match 64.1%; Score 851; DB 22; Length 164;  
 Best Local Similarity 100.0%; Pred. No. 5.6e-80;  
 Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 87 AADDICTAITNGEQVKGILYGPFGTGSFIIGAANOLSKKVRSTIILYPERIRTLKG 146  
 DB 1 AADDICTAITNGEQVKGILYGPFGTGSFIIGAANOLSKKVRSTIILYPERIRTLKG 60  
 QY 147 GFPGDGFSEKILRVREANIIMLDDIGAEVTPWVRDEVIGPLHYRMVHELPTFSSND 206  
 DB 61 GFPGDGFSEKILRVREANIIMLDDIGAEVTPWVRDEVIGPLHYRMVHELPTFSSND 120  
 QY 207 YSELHHLAMTRDGEKTKAARIIRVKSISTPYFLGSENFRRN 250  
 DB 121 YSELHHLAMTRDGEKTKAARIIRVKSISTPYFLGSENFRRN 164

RESULT 6  
 AAB47315  
 ID AAB47315 standard; Protein; 164 AA.  
 AC AAB47315;  
 XX  
 XX 29-AUG-2001 (first entry)

RESULT 7  
 ABB49355  
 ID ABB49355 standard; Protein; 307 AA.  
 AC ABB49355;  
 XX  
 XX 05-FEB-2002 (first entry)  
 DT  
 XX  
 XX Listeria monocytogenes protein #2059.  
 DE  
 XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 KM vitamin B12; bacterial infection; disease.  
 XX



AC ABP27829;  
 XX 02-JUL-2002 (first entry)  
 XX Streptococcus polypeptide SEQ ID NO 4834.  
 DE Streptococcus polypeptide SEQ ID NO 4834.  
 XX Streptococcus GAS; GAS; group B streptococcus; Streptococcus agalactiae;  
 KM group A streptococcus; Streptococcus pyogenes; antibacterial;  
 KM antiinflammatory; infection; vaccine; meningitis; gene therapy.  
 XX Streptococcus pyogenes.  
 OS  
 XX WO200234771-A2.  
 PN 02-MAY-2002.  
 XX 29-OCT-2001; 2001WO-GB04789.  
 XX 27-OCT-2000; 2000GB-0026333.  
 PR 24-NOV-2000; 2000GB-0028727.  
 PR 07-MAR-2001; 2001GB-0005640.  
 XX (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 PI Telford J, Maignani V, Margarit Ros YI, Grandi G, Frazer C;  
 PI Telford J;  
 DR WPI; 2002-352536/38.  
 DR N-PSDB; ABN68460.  
 XX New Streptococcus protein for the treatment or prevention of infection  
 PT or disease caused by Streptococcus bacteria, such as meningitis, and  
 PT for detecting a compound that binds to the protein -  
 XX Claim 1; Page 3645; 4525pp; English.  
 PS The invention relates to a protein (ABP25413-ABP30895) from group B  
 CC streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 CC the specification. The proteins have antibacterial and antiinflammatory  
 CC activity. (1), nucleic acids encoding (1), ABN66044-ABN71526 and  
 CC antibodies that bind (1) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (1) are used to detect Streptococcus in a  
 CC biological sample. (1) is used to determine whether a compound binds to  
 CC (1). A composition comprising (1) or a nucleic acid encoding (1), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (1) may be used to recombinantly produce (1) and may be  
 CC used in gene therapy. Antibodies to (1) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins.  
 XX  
 XX Sequence 300 AA;  
 SQ  
 Query Match 29.9%; Score 397.5; DB 23; Length 300;  
 Best Local Similarity 35.9%; Pred. No. 1.3e-33;  
 Matches 84; Conservative 56; Mismatches 91; Indels 3; Gaps 3;  
 QY 17 PNTV-KHVELYVDNRKIRIKRYIQCCCKIKYDEREALITSHMQRTLANAKLDIY 75  
 DB 68 PSYIAKQYQPLIMNEGAVSVYLETELVEAOKMAISRIOVLSIPKSYRHTLSDID 127  
 QY 76 MNRDRDLVMAADDITAITNGEQVGLVYGPFGKSPFIIGAINOCKSKK-VRSIT 134  
 DB 128 VNAASRKAASALIDFPEQYPSAQ-KGLVLYGMGKSYLLAAMHSEKKGVSTTL 186  
 QY 135 IYVPEFIRTLKGGKSGFEKKLHRYEANILMDDIGAEVTPWVDEVIGPLIHRMV 194  
 DB 187 LHPSPFADVKAMISNOSVVEIDAVNVPEVLILDDIGAGQANISWVDEVLYQVILQYRML 246

QY 195 HELPTFSSNFVDSELEHHLAMTRDGEKTKARILIRVNSTPTPLSGENR 248  
 DB 247 BELPTFTSNYSFADLERKWTATIKGSDETWQAKRVERVYLLAREFLGKANRR 300  
 RESULT 10  
 ID ABP09981 standard; Protein; 92 AA.  
 XX  
 AC ABP09981;  
 XX 25-JUN-2002 (first entry)  
 DE Human ORFX protein sequence SEQ ID NO:19944.  
 XX Human, open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
 KM hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
 KM degenerative disorder; osteoarthritis; neurodegenerative disorder;  
 KM cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
 KM hypertension; hypothyroidism; cholesterol ester storage disease;  
 KM immune deficiency; immune disorder; infectious disease;  
 KM autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
 KM myasthenia gravis.  
 XX  
 OS Homo sapiens.  
 XX WO200192523-A2.  
 PN 06-DEC-2001.  
 XX 29-MAY-2001; 2001WO-US10836.  
 PF 30-MAY-2000; 2000US-206132P.  
 PR 29-AUG-2000; 2000US-228716P.  
 XX (CURA-) CURAGEN CORP.  
 PA  
 PI Shimkets RA, Leach MD;  
 PI WPI; 2002-106308/14.  
 DR N-PSDB; ABN25733.  
 XX Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and autoimmune disorders -  
 XX Disclosure; SEQ ID 19944; 1037pp; English.  
 PS The present invention describes substantially purified human proteins  
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX  
 CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
 CC treating or preventing a pathology associated with an ORFX-associated  
 CC disorder in humans, and in the manufacture of a medicament for treating a  
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
 CC sequences can be used in gene therapy. ORFX sequences can be used in the  
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
 CC storage disease, various immune deficiencies and disorders, infectious  
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
 CC bone degenerative disorders, or periodontal disease, and for gut  
 CC protection or regeneration and treatment of lung or liver fibrosis,  
 CC reperfusion injury in various tissues and conditions resulting from  
 CC systemic cytokine damage.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 92 AA; 29.6%; Score 393; DB 23; Length 92;  
 XX Query Match: Best Local Similarity 81.3%; Pred. No. 7.6e-33; Matches 74; Conservative 10; Mismatches 7; Indels 0; Gaps 0;  
 QY 100 QVAGLVGPGFGKSPILGAINOLKSKKVRSTIYLPFRITLKGFGKGSFEKKLAR 159  
 DB 2 RVAGLTHGPGFGKSPILKSLNOLKSKVRSTIYLPFRITLKGFGKGSFEKKLAR 61  
 QY 160 VREANILMDDIGAEVTPWVRDEVIQPLH 190  
 DB 62 IREANILMDDIGAEVTPWVRDEVIQPLH 92  
 RESULT 11  
 ABP27828  
 ID ABP27828 standard; Protein; 300 AA.  
 AC ABP27828;  
 XX 02-JUL-2002 (first entry)  
 XX Streptococcus polyepide SEQ ID NO 4832.  
 DE Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 XX group A streptococcus; Streptococcus pyogenes; antibacterial;  
 KM antiinflammatory; infection; vaccine; meningitis; gene therapy.  
 XX Streptococcus agalactiae.  
 OS MO200234771-A2.  
 PN 02-MAY-2002.  
 PD 29-OCT-2001; 2001MO-GB04789.  
 PF 27-OCT-2000; 2000GB-0026333.  
 PR 24-NOV-2000; 2000GB-0028727.  
 PR 07-MAR-2001; 2001GB-0005640.  
 XX (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX Telford J, Maignani V, Margart Ros YI, Grandi G, Frazer C;  
 PI Tettein H;  
 XX WPI; 2002-352536/38.  
 DR N-PSDB; ABN68459.  
 XX New Streptococcus protein for the treatment or prevention of infection  
 PT or disease caused by Streptococcus bacteria, such as meningitis, and  
 PT for detecting a compound that binds to the protein -  
 PS Claim 1; Page 3645; 4525pp; English.  
 XX The invention relates to a protein (ABP25413-ABP30895) from group B  
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 CC the specification. The proteins have antibacterial and antiinflammatory  
 CC activity. (1), nucleic acids encoding (1), ABN66044-ABN71526 and  
 CC antibodies that bind (1) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (1) are used to detect Streptococcus in a  
 CC biological sample. (1) is used to determine whether a compound binds to  
 CC (1). A composition comprising (1) or a nucleic acid encoding (1), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (1) may be used to recombinantly produce (1) and may be  
 CC used in gene therapy. Antibodies to (1) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying

CC Streptococcus proteins.  
 XX Sequence 300 AA; 27.5%; Score 365; DB 23; Length 300;  
 XX Query Match: Best Local Similarity 32.4%; Pred. No. 3.1e-29; Matches 77; Conservative 56; Mismatches 103; Indels 2; Gaps 2;  
 QY 12 KPADCNPFKGVPELYVNNRIKIRYLOCPKIKKDEERFAELITSHHMDRLTNAVL 71  
 DB 64 KKKDSQYIANGYEPFLVMWEGYADVSTRELIEHQKQASDRINLVNPSRYNITM 123  
 QY 72 KDIYVNHRLDVLVMAAADICTAITNGEQKGLYYGPGTKSPILGAINOLKSKK-V 130  
 DB 124 TDFDINNESRMKMGQLDFVETPSYNH-KGLIYGDGVGKSYLMAMAEELSERKV 182  
 QY 131 RSTIYLPFRITLKGFGKGSFEKKLARVREANILMDDIGAEVTPWVRDEVIQPLH 190  
 DB 183 STLLHFPSPFALDVKAISSGTVKDEIDAVKSVPLILDDIGAEQATSWVRDEIILQVILQ 242  
 QY 191 YRMVHELPTFPSSNDYSLRHLAMTRDGEKTKARILIERVKSISTEYPLSGENR 248  
 DB 243 HRMLEELPTFTSNYSFNDLERKMANIKGSDETWQKRVMERVRYLAIFFHLEGPNR 300  
 RESULT 12  
 ABU02187  
 ID ABU02187 standard; Protein; 298 AA.  
 XX ABU02187;  
 AC 11-FEB-2003 (first entry)  
 XX S. pneumoniae type 4 strain protein from coding region #1764.  
 DE Bacterial meningitis; pneumonia; sepsis; otitis media;  
 XX ear infection; antiinflammatory; antibacterial; immunostimulant;  
 KM auditory; respiratory; gene therapy; vaccine.  
 XX Streptococcus pneumoniae type 4 strain.  
 OS MO200277021-A2.  
 PN 03-OCT-2002.  
 PD 27-MAR-2002; 2002MO-IB02163.  
 PF 27-MAR-2001; 2001GB-0007658.  
 PR (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX Maignani V, Tettein H, Frazer C;  
 PI Tettein H;  
 XX WPI; 2003-040579/03.  
 DR N-PSDB; ABX07476.  
 XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
 PT useful as medicaments for treating or preventing a disease or infection  
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media  
 PT or ear infection -  
 PS Claim 1; SEQ ID No 3528; 56pp; English.  
 XX The invention relates to a protein comprising or having at least 50%  
 CC identity to any of the 2469 amino acid sequences, identified in the  
 CC specification (available on a computer readable format), or its fragment,  
 CC expressed from 2469 of 2489 identified DNA coding regions from the  
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
 CC ABS35455. Also included are an antibody which binds one of the  
 CC proteins, treating a patient by administering the protein, DNA or  
 CC antibody (in a composition), a kit comprising first and second primers,  
 CC which are the nucleic acid cited above or fragments between nucleotides

Query Match	27.2%	Score 361	DB 24	length 298
Best Local Similarity	34.7%	Pred. No. 8e-29		
Matches 83	Conservative 50	Mismatches 96	Indels 10	Gaps 4

RESULT 13  
ABB54065  
ID ABB54065 standard; Protein; 293 AA.

DT	16-MAY-2002	(first entry)
XX		
DE	Lactococcus lactis	protein dnaI.
XX		
Kw	Biosynthesis, biodegradation, lactic bacterium, yogurt, cheese	
XX		
OS	Lactococcus lactis	IL1403.
XX		
PN	FR2807446-A1.	
XX		
PD	12-OCT-2001.	
XX		
PF	11-APR-2000; 2000FR-0004630.	
XX		
PR	11-APR-2000; 2000FR-0004630.	
XX		
XX		
PA	(INRG ) INRA INST NAT RECH AGRONOMIQUE.	
PI	Bolotine A, Sorokine A, Renault P, Ehrlich SD,	
XX		
RR	WPI; 2002-043418/06.	

Query Match	25.8%	Score 342;	DB 23;	Length 293;
Best Local Similarity	33.3%	Pred. No. 7.3e-27;		
Matches 85;	Conservative 49;	Mismatches 97;	Indels 24;	Gaps 6;

RESULT 14  
ABB53757  
ID ABB53757 standard; Protein; 291 AA

DT	16-MAY-2002 (first entry)
XX	
DE	Lactococcus lactis protein pill4.
XX	
KM	Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
XX	
OS	Lactococcus lactis IL1403.
XX	
FN	FR2807446-A1.
XX	
PD	12-OCT-2001.
XX	
PF	11-APR-2000; 2000FR-0004630.
XX	
PR	11-APR-2000; 2000FR-0004630.
XX	
PA	(INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX	
PI	Bolotine A, Sorokine A, Renault P, Ehrlich SD;
XX	
DR	WPI; 2002-043418/06.
XX	

PT New nucleotide sequence useful in the identification or Lactococcus  
lactic and related species -

PS Claim 6; SEQ ID No 459; 2504bp; French.

XX The present invention is related to a Lactococcus lactic nucleotide  
CC sequence (ABA90521) and related proteins (ABBS3300-ABBS5621). The  
CC nucleic acid sequence is useful in the detection and/or amplification of  
CC nucleic acid sequence, particularly to identify Lactococcus lactic or  
CC related species. The proteins of the invention are useful for the  
CC biosynthesis or biodegradation of a composition of interest. The  
CC invention helps research in lactic bacteria, particularly useful in the  
CC production of yogurt and cheese.  
CC Note: The sequence data for this patent is based on equivalent patent  
CC WO200177334 (published 18-OCT-2001) which is available in electronic  
CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SQ Sequence 291 AA;

Query Match 8.4%; Score 112; DB 23; Length 291;  
Best Local Similarity 21.8%; Pred. No. 0.005; Mismatches 94; Indels 74; Gaps 11;  
Matches 57; Conservative 36; Mismatches 94; Indels 74; Gaps 11;

QY 36 IRYLQCPCKIKYDE---ERFEALITTS-----HMQ-----RDTLN- 68  
DB 29 VAY-KHPVSDIDYDPLHGLVNDGLQVTSYCECAKEGIYQHQQSSAERKANEFLNK 87  
QY 69 -----AKLDIYNNHRDLVMAADICTAITNGEQVGLYL 106  
DB 88 SKYGRSLTKTQSLVGKSLMFPARFNTFKVNGLEQNVLNQORILAREYTOGGRFNTVFV 147  
QY 107 YGPFGRKSFILGALANQ---LKSKKVSTIYLPFRTLKGGKDGSPFKGLHR- 159  
DB 148 -GGAGGKSHLMAVLIQEVNENLKDKFSTLFTNISLREIKNSWNSYSDTKAEERLTT 206  
QY 160 -VEANIILMDDIGAEV---TPWVDEVIGPLHYRMWHELPFSSNFDYSELEHLL 214  
DB 207 LKRTYDVLVIDDGTSTSTSKNSWQ---GVYVNIYARBSNTIITSNLTKEM- 258  
QY 215 AMTRDGEKTKAARIIRYKS 235  
DB 259 ---RSSYDDKIYSRIWEGSKN 276

RESULT 15

ABBS4759  
ID ABBS4759 standard; Protein; 294 AA.

AC ABBS4759;

DT 16-MAY-2002 (first entry)

DE Lactococcus lactic protein p1346.

KM Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.

OS Lactococcus lactic IL1403.

PN FR2807446-A1.

PD 12-OCT-2001.

PF 11-APR-2000; 2000FR-0004630.

PR 11-APR-2000; 2000FR-0004630.

PA (INRG ) INRA INST NAT RECH AGRONOMIQUE.

PI Bolocrine A, Sorokine A, Renault P, Ehrlich SD;

WPI; 2002-043418/06.

PT New nucleotide sequence useful in the identification or Lactococcus

PT lactic and related species -

PS Claim 6; SEQ ID No 1461; 2504bp; French.

XX The present invention is related to a Lactococcus lactic nucleotide  
CC sequence (ABA90521) and related proteins (ABBS3300-ABBS5621). The  
CC nucleic acid sequence is useful in the detection and/or amplification of  
CC nucleic acid sequence, particularly to identify Lactococcus lactic or  
CC related species. The proteins of the invention are useful for the  
CC biosynthesis or biodegradation of a composition of interest. The  
CC invention helps research in lactic bacteria, particularly useful in the  
CC production of yogurt and cheese.  
CC Note: The sequence data for this patent is based on equivalent patent  
CC WO200177334 (published 18-OCT-2001) which is available in electronic  
CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SQ Sequence 294 AA;

Query Match 8.2%; Score 109.5; DB 23; Length 294;  
Best Local Similarity 22.3%; Pred. No. 0.0092; Mismatches 88; Indels 55; Gaps 8;  
Matches 50; Conservative 31; Mismatches 88; Indels 55; Gaps 8;

QY 55 ELTSHMQ-----RDTLN-----AKLDIYNNHRDL 83  
DB 68 EGIYQHQQSSAERKANEFLNKSKYKSLTKQSLVGKSLMFPARFNTFKNSLEQN 127  
QY 84 VMAADICTAITNGEQVGLYLPFGRKSFILGALANQ---LKSKKVSTIYLPF 139  
DB 128 VLNQORILAREYTOGGRFNTVFV-GGAGGKSHLMAVLIQEVNENLKDKFSTLFTNISB 186  
QY 140 FIRTLLKGGKDGSPFKGLHR---VEANIILMDDIGAEV---TPWVDEVIGPLHY 191  
DB 187 LIREIKNSWNSYSDTKAEERLTTMKRTYDVLVIDDGTSTSTSKNSWQ---GVYVNI 242  
QY 192 RMVHELPTFSSNFDYSELEHLLAMTRDGEKTKAARIIRYKS 235  
DB 243 VNAREGNTIITSNLTKEM-----RSSYDDKIYSRIWEGSKN 279

Search completed: December 15, 2003, 15:13:43  
Job time : 36.4512 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: December 15, 2003, 15:11:41 ; Search time 13.0674 Seconds  
(without alignments)  
809.474 Million cell updates/sec

Title: US-09-689-952-18

Sequence: 1 YDDQKHVDGKHFADCPNRY.....EVKSLSTYPLSGENFRNN 250

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:  
1: /cgnt\_6/prodata/1/aa/5A\_COMB.pep:\*  
2: /cgnt\_6/prodata/1/aa/5B\_COMB.pep:\*  
3: /cgnt\_6/prodata/1/aa/5A\_COMB.pep:\*  
4: /cgnt\_6/prodata/1/aa/5B\_COMB.pep:\*  
5: /cgnt\_6/prodata/1/aa/5A\_COMB.pep:\*  
6: /cgnt\_6/prodata/1/aa/5B\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1328	100.0	313	4	US-09-470-512A-2
2	1315	99.0	313	4	US-09-470-512A-10
3	1189	89.5	307	4	US-09-134-001C-3792
4	537.5	40.5	280	4	US-09-470-512A-14
5	531.5	40.0	278	4	US-09-470-512A-15
6	354	26.7	310	4	US-09-107-532A-5580
7	124	9.3	268	4	US-09-328-352-5963
8	102.5	7.7	304	4	US-09-134-001C-5153
9	92.5	7.0	467	4	US-09-198-452A-443
10	90	6.8	389	2	US-08-820-170A-13
11	90	6.8	389	3	US-09-055-699-13
12	90	6.8	389	3	US-09-273-565-13
13	90	6.8	389	4	US-09-565-538-13
14	90	6.8	389	4	US-09-661-468-13
15	90	6.8	389	4	US-09-976-165-13
16	89.5	6.7	717	4	US-08-924-629C-5
17	89	6.7	1349	2	US-08-612-734B-2
18	88	6.6	1334	2	US-08-996-545-2
19	88	6.6	1334	3	US-09-328-320-2
20	88	6.6	1438	4	US-09-209-916-1
21	87.5	6.6	376	4	US-09-252-991A-24606
22	87.5	6.6	567	4	US-09-134-001C-5646
23	87	6.6	872	4	US-09-198-452A-163
24	86	6.5	468	4	US-09-134-001C-5156
25	86	6.5	529	4	US-09-252-991A-18830
26	86	6.5	957	4	US-09-107-532A-4021
27	85.5	6.4	928	2	US-08-841-483-4

28	85.5	6.4	928	3	US-09-382-911-4	Sequence 4, Appl1
29	85.5	6.4	1117	2	US-08-841-483-6	Sequence 6, Appl1
30	85.5	6.4	1117	3	US-09-382-911-6	Sequence 6, Appl1
31	84.5	6.4	527	4	US-09-107-532A-4822	Sequence 4822, Ap
32	84.5	6.4	795	4	US-09-328-352-6143	Sequence 6143, Ap
33	83.5	6.3	266	4	US-09-218-363-8	Sequence 8, Appl1
34	83.5	6.3	267	4	US-09-218-363-10	Sequence 10, Appl1
35	83.5	6.3	290	4	US-09-218-363-4	Sequence 4, Appl1
36	83.5	6.3	456	3	US-09-268-364-21	Sequence 21, Appl1
37	83	6.2	308	4	US-09-149-476-607	Sequence 607, App
38	83	6.2	399	4	US-09-149-476-474	Sequence 474, App
39	83	6.2	806	3	US-08-999-774A-6	Sequence 6, Appl1
40	82.5	6.2	427	3	US-09-196-857-2	Sequence 2, Appl1
41	82.5	6.2	1964	2	US-08-790-912-3	Sequence 3, Appl1
42	82.5	6.2	2052	2	US-08-790-912-2	Sequence 2, Appl1
43	82	6.2	227	4	US-09-291-170A-9	Sequence 9, Appl1
44	82	6.2	227	4	US-09-724-884-9	Sequence 9, Appl1
45	82	6.2	1471	1	US-08-683-839B-3	Sequence 3, Appl1

## ALIGNMENTS

```

RESULT 1
US-09-470-512A-2
; Sequence 2, Application US/09470512A
; Patent No. 6376652
; GENERAL INFORMATION:
; APPLICANT: Phagotech, Inc.
; TITLE OF INVENTION: Compositions and methods involving an essential Staphylococcus Au
; TITLE OF INVENTION: gene and its encoded protein
; FILE REFERENCE: 21715/1000
; CURRENT APPLICATION NUMBER: US/09/470,512A
; CURRENT FILING DATE: 1999-12-12
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 313
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-470-512A-2

Query Match      100.0%; Score 1328; DB 4; Length 313;
Best Local Similarity 100.0%; Pred. No. 2.5e-141;
Matches 250; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 YDDQKHVDGKHFADCPNRYKGVHVELYDNNRIKIRYLCQCKIKYDERFEALITSH 60
DB      64 YDDQKHVDGKHFADCPNRYKGVHVELYDNNRIKIRYLCQCKIKYDERFEALITSH 123
QY      61 HMQDPTLNKLDIYNNHRDLVDVMAADDICTAITNGEQVGLYVYGFPGTKSFILGA 120
DB      124 HMQDPTLNKLDIYNNHRDLVDVMAADDICTAITNGEQVGLYVYGFPGTKSFILGA 183
QY      121 INQKSKVRSRTIYLPFRITLKGFGDGFSEKKLHVRANILMLDDIGAEVTPMV 180
DB      184 INQKSKVRSRTIYLPFRITLKGFGDGFSEKKLHVRANILMLDDIGAEVTPMV 243
QY      181 RDEVIGPLAHYMHVELPFFSSNFDYSELHHLAMTRDGEKTKAARIIEVKSLSTY 240
DB      244 RDEVIGPLAHYMHVELPFFSSNFDYSELHHLAMTRDGEKTKAARIIEVKSLSTY 303
QY      241 FLSGENFRNN 250
DB      304 FLSGENFRNN 313

RESULT 2
US-09-470-512A-10
; Sequence 10, Application US/09470512A
; Patent No. 6376652
; GENERAL INFORMATION:
; APPLICANT: Phagotech, Inc.

```

;; TITLE OF INVENTION: Compositions and methods involving an essential Staphylococcus Au  
;; FILE OF INVENTION: gene and its encoded protein  
;; FILE REFERENCE: 21715/1000  
;; CURRENT APPLICATION NUMBER: US/09/470,512A  
;; CURRENT FILING DATE: 1999-12-12  
;; NUMBER OF SEQ ID NOS: 16  
;; SOFTWARE: PatentIn version 3.0  
;; SEQ ID NO 10  
;; LENGTH: 313  
;; TYPE: PRF  
;; ORGANISM: Staphylococcus aureus  
US-09-470-512A-10

Query Match 99.0%; Score 1315; DB 4; Length 313;  
Best Local Similarity 98.4%; Pred. No. 7,3e-140;  
Matches 246; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 YKQOQKHVDGKADCPNFKGVHPELVYDNNRIKIRYLQCPCKIKYDEERFEALITSH 60  
DB 64 YKQOQKHVDGKADCPNFKGVHPELVYDNNRIKIRYLQCPCKIKYDEERFEALITSH 123  
QY 61 HMQORDTLNAKLDIYNNHRLDVMAADICTAINTGEQVKGILYGPFGTGSFLGA 120  
DB 124 HMQORDTLNAKLDIYNNHRLDVMAADICTAINTGEQVKGILYGPFGTGSFLGA 183  
QY 121 IANQLSKKVRSITIIYPERIRTLKGFGKGSFEKTLHVRREANIIMLDIGAEVTPWV 180  
DB 184 IANQLSKKVRSITIIYPERIRTLKGFGKGSFEKTLHVRREANIIMLDIGAEVTPWV 243  
QY 181 RDEVIGPLHYRMVHELPTFSSNFDYSELEHLLAMTRDGEETKARIIERYKSLSTPY 240  
DB 244 RDEVIGPLHYRMVHELPTFSSNFDYSELEHLLAMTRDGEETKARIIERYKSLSTPY 303  
QY 241 FLGSENFRRN 250  
DB 304 FLGSENFRRN 313

## RESULT 3

US-09-134-001C-3792  
;; Sequence 3792, Application US/09134001C  
;; Patent No. 6380370  
;; GENERAL INFORMATION:  
;; APPLICANT: Lyta Doucette-Stamm et al  
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
;; TITLE OF INVENTION: EPIDERMIDS FOR DIAGNOSTICS AND THERAPEUTICS  
;; FILE REFERENCE: GTC-007  
;; CURRENT APPLICATION NUMBER: US/09/134,001C  
;; CURRENT FILING DATE: 1998-08-13  
;; PRIOR APPLICATION NUMBER: US 60/064,964  
;; PRIOR FILING DATE: 1997-11-08  
;; PRIOR APPLICATION NUMBER: US 60/055,779  
;; PRIOR FILING DATE: 1997-08-14  
;; NUMBER OF SEQ ID NOS: 5674  
;; SEQ ID NO 3792  
;; LENGTH: 307  
;; TYPE: PRF  
;; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3792

Query Match 89.5%; Score 1189; DB 4; Length 307;  
Best Local Similarity 87.2%; Pred. No. 1.1e-125;  
Matches 218; Conservative 23; Mismatches 9; Indels 0; Gaps 0;

QY 1 YKQOQKHVDGKADCPNFKGVHPELVYDNNRIKIRYLQCPCKIKYDEERFEALITSH 60  
DB 58 YKQOQKHVDGKADCPNFKGVHPELVYDNNRIKIRYLQCPCKIKYDEERFEALITSH 117  
QY 61 HMQORDTLNAKLDIYNNHRLDVMAADICTAINTGEQVKGILYGPFGTGSFLGA 120  
DB 118 HMQORDTLNAKLDIYNNHRLDVMAADICTAINTGEQVKGILYGPFGTGSFLGA 177  
QY 121 IANQLSKKVRSITIIYPERIRTLKGFGKGSFEKTLHVRREANIIMLDIGAEVTPWV 180

DB 178 IANQLSKKVRSITIIYPERIRTLKGFGKGSFEKTLHVRREANIIMLDIGAEVTPWV 237  
QY 181 RDEVIGPLHYRMVHELPTFSSNFDYSELEHLLAMTRDGEETKARIIERYKSLSTPY 240  
DB 238 RDEVIGPLHYRMVHELPTFSSNFDYSELEHLLAMTRDGEETKARIIERYKSLSTPY 297  
QY 241 FLGSENFRRN 250  
DB 298 YLTGKNFRRN 307

## RESULT 4

US-09-470-512A-14  
;; Sequence 14, Application US/09470512A  
;; Patent No. 6376652  
;; GENERAL INFORMATION:  
;; APPLICANT: PhageTech, Inc.  
;; TITLE OF INVENTION: Compositions and methods involving an essential Staphylococcus Au  
;; TITLE OF INVENTION: gene and its encoded protein  
;; FILE REFERENCE: 21715/1000  
;; CURRENT APPLICATION NUMBER: US/09/470,512A  
;; CURRENT FILING DATE: 1999-12-12  
;; NUMBER OF SEQ ID NOS: 16  
;; SOFTWARE: PatentIn version 3.0  
;; SEQ ID NO 14  
;; LENGTH: 280  
;; TYPE: PRF  
;; ORGANISM: Bacillus subtilis  
US-09-470-512A-14

Query Match 40.5%; Score 537.5; DB 4; Length 280;  
Best Local Similarity 44.9%; Pred. No. 2.8e-52;  
Matches 106; Conservative 48; Mismatches 81; Indels 1; Gaps 1;

QY 15 DCPNFKGVHPELVYDNNRIKIRYLQCPCKIKYDEERFEALITSHHMQORDTLNAKLDI 74  
DB 45 NCMNLEGVHPLVNGRSIDIEYECVYKRLDQKQGSIMKSMYIQDILGATFGQV 104  
QY 75 YNNHRLDVMAADICTAINTGEQVKGILYGPFGTGSFLGAIANQLSKKVRSIT 134  
DB 105 DISDPSRLMFGVHVDPLFKSYNETGKGLYLGKGVGTPTMLAIANLEKEYSSMI 164  
QY 135 IYLPERRITLKGFGKGSFEKTLHVRREANIIMLDIGAEVTPWVRDEVIGPLHYRMV 194  
DB 165 VYPERVRLKSLSDQITSEKLNWKTTPVIMLDIGAESVTSNVRDEVITGLQHRNS 224  
QY 195 HELPTFSSNFDYSELEHLLAMTRDGE-EKTKAARIERYKSLSTPYFLGSENFRRN 249  
DB 225 QCLPTFSSNFDYSELEHLLAMTRDGE-EKTKAARIERYKSLSTPYFLGSENFRRN 280

## RESULT 5

US-09-470-512A-15  
;; Sequence 15, Application US/09470512A  
;; Patent No. 6376652  
;; GENERAL INFORMATION:  
;; APPLICANT: PhageTech, Inc.  
;; TITLE OF INVENTION: Compositions and methods involving an essential Staphylococcus Au  
;; TITLE OF INVENTION: gene and its encoded protein  
;; FILE REFERENCE: 21715/1000  
;; CURRENT APPLICATION NUMBER: US/09/470,512A  
;; CURRENT FILING DATE: 1999-12-12  
;; NUMBER OF SEQ ID NOS: 16  
;; SOFTWARE: PatentIn version 3.0  
;; SEQ ID NO 15  
;; LENGTH: 278  
;; TYPE: PRF  
;; ORGANISM: Bacillus halodurans  
US-09-470-512A-15

Query Match 40.0%; Score 531.5; DB 4; Length 278;  
Best Local Similarity 41.2%; Pred. No. 1.3e-51;

RESULT 6  
 US-09-107-532A-5580  
 Sequence 5580, Application US/09107532A  
 Patent No. 5583275  
 GENERAL INFORMATION:  
 APPLICANT: Lynn A Doucette-Stamm and David Bush  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
 ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
 NUMBER OF SEQUENCES: 7310  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
 STREET: 100 Beaver Street  
 CITY: Waltham  
 STATE: Massachusetts  
 COUNTRY: USA  
 ZIP: 02354  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: CD-ROM ISO9660  
 COMPUTER: PC  
 OPERATING SYSTEM: <Unknown>  
 SOFTWARE: ASCII  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/107.532A  
 FILING DATE: 30-Jun-1998  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 60/085,598  
 FILING DATE: 14 May 1998  
 APPLICATION NUMBER: 60/051571  
 FILING DATE: July 2, 1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Arinello, Pamela Deneke  
 REGISTRATION NUMBER: 40,489  
 REFERENCE/DOCKET NUMBER: GTC-012  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (781)893-5007  
 TELEFAX: (781)893-8277  
 INFORMATION FOR SEQ ID NO: 5580:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 310 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: YES  
 ORIGINAL SOURCE:  
 ORGANISM: Enterococcus faecium  
 FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: (B) LOCATION 1...310  
 SEQUENCE DESCRIPTION: SEQ ID NO: 5580:  
 IS-09-107-532A-5580

```

RESULT 7
US-09-328-352-5963
; Sequence 5963, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5963
; LENGTH: 268
; TYPE: PR1
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5963

Query Match          9.3%; Score 124; DB 4; Length 268;
Best Local Similarity 25.6%; Pred. No. 1.1e-05;
Matches 34; Conservative 28; Mismatches 63; Indels 8; Gaps 4;

Cy      44 KIKYDEEFEEELTYTSHMQRDITLAKLK-DIYNHRRRLDVAMADICTAIT--NGEQ 100
       |||::||::||::||::||::||::||::||::||::||::||::||::||
Db      102 KVKEPEPVANATVASSSSQKKLLNPQFTFLFVEGRN---QMAETRKVLTLGASQ 158

Cy      101 VKGLYLGPFGTGKSFILGAIANOLKSKKVRSITYL--PEFIRFLKGSGFKDGSFEKKLH 158
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      159 HNPFLYPTGTGLQKTHLMQAIVGNALLQAKPNARVMVTWTSQSFFVDQFVSLSIQKGYEERFK 218

Cy      159 RVRANITMLDDI 171
       |:::|||
Db      219 NCRLDLILLVDI 231

RESULT 8
US-09-134-001C-5153
; Sequence 5153, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OR INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674

```

SEQ ID NO 5153  
LENGTH: 304  
TYPE: PRF  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5153

Query Match  
Best Local Similarity 21.4%; Score 102.5; DB 4; Length 304;  
Best Local Similarity 21.4%; Pred. No. 0.0035;  
Matches 33; Conservative 35; Mismatches 59; Indels 27; Gaps 4;

QY 88 ADDICATINNGEBOVKGLYGPPTGKSFILGAIANOLKSKVRSSTIYLP-----EPI 141  
DB 135 AAGLAVALAPAEKAVNPFIYGVGLGKTHMAIGHVLSNKPNAKVIYTSSEKFTWEFI 194  
QY 142 RTKAGFKDGSFPEKHLRVREANILMDIGAEVTPWVDEVYIGPLHVRMHELPTF 201  
DB 195 KSI-----RDNETAFREKRYKIVLLDIOPTONKEQTGE-----FF 234  
QY 202 SSNPDYSELEHHLAMTDEGEKTKAARIRKRS 235  
DB 235 HTFNEHLQNNKQIVISSDRPK-EIAKLEDRRS 267

RESULT 9  
US-09-198-452A-443  
Sequence 443, Application US/09198452A  
Patent No. 6559294  
GENERAL INFORMATION:  
APPLICANT: Griffiths, R.  
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments  
TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention  
FILE REFERENCE: 9710-003-999  
CURRENT APPLICATION NUMBER: US/09/198,452A  
CURRENT FILING DATE: 1998-11-24  
NUMBER OF SEQ ID NOS: 6849  
SEQ ID NO 443  
LENGTH: 467  
TYPE: PRF  
ORGANISM: Chlamydia pneumoniae  
US-09-198-452A-443

Query Match  
Best Local Similarity 21.0%; Score 92.5; DB 4; Length 467;  
Best Local Similarity 21.0%; Pred. No. 0.092;  
Matches 38; Conservative 39; Mismatches 73; Indels 31; Gaps 8;

QY 20 VKGVELVYDNNRIKIRYLQCPCK-ICKDEERFEA-ELITSHR-----MORDTLNA 69  
DB 60 IRLEVENIIFQNYLID-NYKRDLCSEFPLDVHGEPALFEVVAHKKPSAPVASQKES-NE 117  
QY 70 KKDITMNRD---RDLVMAADDICATIN-----GEQVKGILYGPPTG 112  
DB 118 GISEVFEETDFELKLNLSYRPNFLEGPSNQVKSAAVGIACKPGSYNPLFIHGCVGL 177  
QY 113 GKSFILGAIANOLKSKVRSSTIYLPFIYGVGLGKTHMAIGHVLSNKPNAKVIYTSSEKFTWEFI 170  
DB 178 GKTHLHVAHVYREHKKLRICITTEATINDLVHLKSKVDKMKNFYRSDDLVD 237

QY 171 I 171  
DB 238 I 238

RESULT 10  
US-08-820-170A-13  
Sequence 13, Application US/08820170A  
Patent No. 5831058  
GENERAL INFORMATION:  
APPLICANT: Tsutomu, FUJIWARA  
APPLICANT: Takeshi, WATANABE  
APPLICANT: Masato, HORIE  
APPLICANT: Toyomasa, KATAGIRI  
TITLE OF INVENTION: HUMAN GENE

NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Sughrue, Mion, Zinn, Macpeak & Seas  
STREET: 2100 Pennsylvania Avenue, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: United States  
ZIP: 20037-3202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/820,170A  
FILING DATE:  
CLASSIFICATION: 536  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 293-7060  
TELEFAX: (202) 293-7860  
TELEX: 6491103  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 389 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-820-170A-13

Query Match  
Best Local Similarity 35.1%; Score 90; DB 2; Length 389;  
Best Local Similarity 35.1%; Pred. No. 0.13;  
Matches 27; Conservative 13; Mismatches 25; Indels 12; Gaps 3;

QY 102 KGLIYGPPTGKSFILGAIANOLKSKVRSSTIYLPFIYGVGLGKTHMAIGHVLSNKPNAKVIYTSSEKFTWEFI 156  
DB 168 KGCLYGPPTGKSFILGAIANOLKSKVRSSTIYLPFIYGVGLGKTHMAIGHVLSNKPNAKVIYTSSEKFTWEFI 220  
QY 157 LHRVREANILMDIGA 173  
DB 221 YARDHQPCLIFMDEIDA 237

RESULT 11  
US-09-055-699-13  
Sequence 13, Application US/09055699  
Patent No. 6005088  
GENERAL INFORMATION:  
APPLICANT: Tsutomu, FUJIWARA  
APPLICANT: Takeshi, WATANABE  
APPLICANT: Masato, HORIE  
APPLICANT: Toyomasa, KATAGIRI  
TITLE OF INVENTION: HUMAN GENE  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Sughrue, Mion, Zinn, Macpeak & Seas  
STREET: 2100 Pennsylvania Avenue, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: United States  
ZIP: 20037-3202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/055,699  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/820,170  
FILING DATE:  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 293-7060  
TELEFAX: (202) 293-7860  
TELEX: 6491103  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 389 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-055-699-13

Query Match 6.8%; Score 90; DB 3; Length 389;  
Best Local Similarity 35.1%; Pred. No. 0.13;  
Matches 27; Conservative 13; Mismatches 25; Indels 12; Gaps 3;

QY 102 KGLVYGPFGTGSFLGAIANQLKS--KQVRSITL--YLPEFIRTLKGFQDGSFEKK 156  
DB 168 KGCCLYGPFGTGTLLARAVASQLDCNFKLVSSIVDKYIGESARLLIREMF-----N 220  
QY 157 LHRVREANILMDDIGA 173  
DB 221 YARDHQPCTIFMDEIDA 237

RESULT 12  
US-09-273-565-13  
Sequence 13, Application US/09273565A  
Patent No. 6166190

GENERAL INFORMATION:  
APPLICANT: FUJIMURA, TSUTOMU  
APPLICANT: WATANABE, TAKESHI  
APPLICANT: HORIE, MASATO  
TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN  
FILE REFERENCE: Q-53599  
CURRENT APPLICATION NUMBER: US/09/273,565A  
EARLIER FILING DATE: 1999-03-22  
EARLIER APPLICATION NUMBER: 09/055,699  
EARLIER FILING DATE: 1998-04-07  
EARLIER APPLICATION NUMBER: 08/820,170  
EARLIER FILING DATE: 1997-03-19  
EARLIER APPLICATION NUMBER: JP 63410/1996  
EARLIER FILING DATE: 1996-03-19  
EARLIER APPLICATION NUMBER: JP 69163/1997  
EARLIER FILING DATE: 1997-03-05  
NUMBER OF SEQ ID NOS: 95  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 13  
LENGTH: 389  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-273-565-13

Query Match 6.8%; Score 90; DB 3; Length 389;  
Best Local Similarity 35.1%; Pred. No. 0.13;  
Matches 27; Conservative 13; Mismatches 25; Indels 12; Gaps 3;

QY 102 KGLVYGPFGTGSFLGAIANQLKS--KQVRSITL--YLPEFIRTLKGFQDGSFEKK 156  
DB 168 KGCCLYGPFGTGTLLARAVASQLDCNFKLVSSIVDKYIGESARLLIREMF-----N 220  
QY 157 LHRVREANILMDDIGA 173  
DB 221 YARDHQPCTIFMDEIDA 237

RESULT 13  
US-09-565-538-13  
Sequence 13, Application US/09565538  
Patent No. 6333404  
GENERAL INFORMATION:  
APPLICANT: FUJIMURA, TSUTOMU  
APPLICANT: WATANABE, TAKESHI

APPLICANT: HORIE, MASATO  
TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN  
FILE REFERENCE: Q-53599  
CURRENT APPLICATION NUMBER: US/09/565,538  
CURRENT FILING DATE: 2000-05-05  
PRIOR APPLICATION NUMBER: 09/273,565  
PRIOR FILING DATE: 1999-03-22  
PRIOR APPLICATION NUMBER: 09/055,699  
PRIOR FILING DATE: 1998-04-07  
PRIOR APPLICATION NUMBER: 08/820,170  
PRIOR FILING DATE: 1997-03-19  
PRIOR APPLICATION NUMBER: JP 63410/1996  
PRIOR FILING DATE: 1996-03-19  
PRIOR APPLICATION NUMBER: JP 69163/1997  
NUMBER OF SEQ ID NOS: 95  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 13  
LENGTH: 389  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-565-538-13

Query Match 6.8%; Score 90; DB 4; Length 389;  
Best Local Similarity 35.1%; Pred. No. 0.13;  
Matches 27; Conservative 13; Mismatches 25; Indels 12; Gaps 3;

QY 102 KGLVYGPFGTGSFLGAIANQLKS--KQVRSITL--YLPEFIRTLKGFQDGSFEKK 156  
DB 168 KGCCLYGPFGTGTLLARAVASQLDCNFKLVSSIVDKYIGESARLLIREMF-----N 220  
QY 157 LHRVREANILMDDIGA 173  
DB 221 YARDHQPCTIFMDEIDA 237

RESULT 14  
US-09-661-468-13  
Sequence 13, Application US/09661468  
Patent No. 6376189

GENERAL INFORMATION:  
APPLICANT: FUJIMURA, TSUTOMU  
APPLICANT: WATANABE, TAKESHI  
APPLICANT: HORIE, MASATO  
TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN  
FILE REFERENCE: Q-53599  
CURRENT APPLICATION NUMBER: US/09/661,468  
CURRENT FILING DATE: 2000-09-13  
PRIOR APPLICATION NUMBER: 09/055,699  
PRIOR FILING DATE: 1998-04-07  
PRIOR APPLICATION NUMBER: 08/820,170  
PRIOR FILING DATE: 1997-03-19  
PRIOR APPLICATION NUMBER: JP 63410/1996  
PRIOR FILING DATE: 1996-03-19  
PRIOR APPLICATION NUMBER: JP 69163/1997  
NUMBER OF SEQ ID NOS: 95  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 13  
LENGTH: 389  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-661-468-13

Query Match 6.8%; Score 90; DB 4; Length 389;  
Best Local Similarity 35.1%; Pred. No. 0.13;  
Matches 27; Conservative 13; Mismatches 25; Indels 12; Gaps 3;

QY 102 KGLVYGPFGTGSFLGAIANQLKS--KQVRSITL--YLPEFIRTLKGFQDGSFEKK 156  
DB 168 KGCCLYGPFGTGTLLARAVASQLDCNFKLVSSIVDKYIGESARLLIREMF-----N 220

```
Qy      157 LHRVREANILMLDDIGA 173
      | : | : | |
Db      221 YARDHQPCITFMEIDA 237
```

**RESULT 15**

```

US-09-976-165-13
Sequence 13. Application US/09976165
Patent No. 6562947
GENERAL INFORMATION:
APPLICANT: FUJIMARA, TSUTOMU
APPLICANT: MATANABE, TAKESHI
APPLICANT: HORIE, MASATO
TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
FILE REFERENCE: Q-53599
CURRENT APPLICATION NUMBER: US/09/976,165
CURRENT FILING DATE: 2001-10-15
PRIORITY APPLICATION NUMBER: 09/565,538
PRIORITY FILING DATE: 2000-05-05
PRIORITY APPLICATION NUMBER: 09/055,699
PRIORITY FILING DATE: 1998-04-07
PRIORITY APPLICATION NUMBER: 08/820,170
PRIORITY FILING DATE: 1997-03-19
PRIORITY APPLICATION NUMBER: JP 62410/1996
PRIORITY FILING DATE: 1996-03-19
PRIORITY APPLICATION NUMBER: JP 62163/1997
PRIORITY FILING DATE: 1997-03-05
NUMBER OF SEQ ID NOS: 95
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 13
LENGTH: 389
TYPE: PRT
ORGANISM: Homo sapiens
US-09-976-165-13

```

Query Match 6.8%; Score 90; DB 4; Length 389;  
 Best Local Similarity 35.1%; Pred. No. 0.13;  
 Matches 27; Conservative 13; Mismatches 12; Gaps 3;

Qy 102 KGLVLYGPFGTGSFLIGAIANOLKS---KVRSTII--YPEFIRTLKGGFKDGSPEKK 156  
Db 168 KGLGLVYGPFGTGTTLARAVASQDLQCNFLKVVSSSIYDKYIGESARLRNEMF-----N 220

Qy 157 LHRVRBANILMLDDIGA 173  
Db 221 YARDHQPCTIIFMEDEIDA 237

```
Search completed: December 15, 2003, 15:17:40
Job time : 14.0674 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 15:13:51 / Search time 24.4154 Seconds  
(without alignments)  
1904.368 Million cell updates/sec

Title: US-09-689-952-18  
Perfect score: 1328  
Sequence: 1 YDQCKHYDKHFKADCPNFV.....ERVKSISTPYLSEGFRRNN 250

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues  
Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications AA.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	7.9	248	12	US-10-287-274-435
2	100	7.5	463	9	US-10-156-761-11850
3	98	7.4	657	9	US-09-815-242-10395
4	96.5	7.3	442	9	US-09-815-242-13552
5	96	7.2	454	9	US-09-815-242-11167
6	93.5	7.0	796	12	US-10-032-585-7201
7	92.5	7.0	574	10	US-09-764-868-774
8	91.5	6.9	827	12	US-10-032-585-7621
9	90.5	6.8	577	9	US-09-815-242-10193
10	90	6.8	389	10	US-09-976-165-113
11	90	6.8	389	12	US-10-205-219-165
12	90	6.8	389	12	US-10-342-276-113
13	90	6.8	393	15	US-10-128-714-3421
14	90	6.8	445	9	US-09-815-242-5010
15	90	6.8	447	9	US-09-815-242-10908

16	90	6.8	451	15	US-10-128-714-8421	Sequence 8421, App
17	89.5	6.7	717	11	US-09-883-343A-5	Sequence 5, Appl1
18	88	6.6	444	12	US-10-032-585-7385	Sequence 7385, App
19	88	6.6	453	9	US-09-815-242-5668	Sequence 5668, App
20	88	6.6	453	9	US-09-815-242-12701	Sequence 12701, App
21	88	6.6	1334	9	US-09-758-828-2	Sequence 2, Appl1
22	88	6.6	1338	14	US-10-006-091-1	Sequence 1, Appl1
23	88	6.6	1438	14	US-10-047-257-1	Sequence 1, Appl1
24	88	6.6	1438	15	US-10-225-960-1	Sequence 1, Appl1
25	87.5	6.6	411	12	US-10-032-585-7284	Sequence 7284, App
26	87	6.6	682	12	US-10-032-585-7047	Sequence 7047, App
27	87	6.6	866	12	US-09-841-260-113	Sequence 113, App
28	87	6.6	866	14	US-10-007-693-113	Sequence 113, App
29	86.5	6.5	606	15	US-10-128-714-3064	Sequence 3064, App
30	86.5	6.5	847	15	US-10-128-714-8064	Sequence 8064, App
31	86	6.5	439	15	US-10-128-714-3522	Sequence 3522, App
32	86	6.5	513	15	US-10-128-714-8522	Sequence 8522, App
33	86	6.5	439	15	US-09-815-245-11772	Sequence 11772, App
34	85.5	6.4	298	12	US-10-380-558-17	Sequence 17, Appl1
35	85.5	6.4	444	15	US-10-153-668-308	Sequence 308, App
36	85.5	6.4	465	9	US-09-925-302-771	Sequence 771, App
37	85.5	6.4	482	10	US-09-764-868-1149	Sequence 1149, App
38	85.5	6.4	489	10	US-09-764-868-776	Sequence 776, App
39	85	6.4	421	15	US-10-128-714-8485	Sequence 8485, App
40	85	6.4	465	15	US-10-128-714-3485	Sequence 3485, App
41	84.5	6.4	442	11	US-09-828-062-7	Sequence 7, Appl1
42	84.5	6.4	584	10	US-09-764-868-1054	Sequence 1054, App
43	84	6.3	401	12	US-10-032-585-7325	Sequence 7325, App
44	84	6.3	508	11	US-09-769-787-102	Sequence 102, App
45	84	6.3	561	15	US-10-156-761-13374	Sequence 13374, App

#### ALIGNMENTS

RESULT 1  
US-10-287-274-435  
; Sequence 435, Application US/10287274  
; Publication No. US20030181408A1  
; GENERAL INFORMATION:  
; APPLICANT: Forsyth, R. Allyn  
; APPLICANT: Ohlsen, Karl  
; APPLICANT: Zyskind, Judith  
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERETO  
; FILE REFERENCE: ELIIRA, 008DVI  
; CURRENT APPLICATION NUMBER: US/10/287, 274  
; PRIOR FILING DATE: 2002-10-31  
; PRIOR APPLICATION NUMBER: US 60/164415  
; PRIOR FILING DATE: 1999-11-09  
; PRIOR APPLICATION NUMBER: US 09/111164  
; NUMBER OF SEQ ID NOS: 469  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 435  
; LENGTH: 248  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-10-287-274-435  
Query Match 7.9%; Score 105; DB 12; Length 248;  
Best Local Similarity 26.7%; Pred. No. 0.004;  
Matches 43; Conservative 30; Mismatches 66; Indels 22; Gaps 7;  
QY 77 NHRDLVAMA---ADICITATNGEQVGLYLPGFGTGSFIIAGINOLKSKVRBT 133  
DB 80 NNGCQYALSAKSIDELMTGCTN-----FAFGSGPGTGKXHLAALINRLKQGTVI 133  
QY 134 IYLBEPFRTLGKRGDQ-SPEKKLHYREANIIMDDIGAEVTPWVDEYIGPLHYR 192  
DB 134 VYTVADVWSALASTDDGSGKFLRELCEVDLVDELGIQRET---KNEQV--VLH-Q 187  
QY 193 WYHELTPFSSNPDYSELEHHILAMTRDGEKTKAARIERV 233

Db 188 IVDRTASMSVGLTNVLEAMKTLGGE-----RINDRM 222

RESULT 2

US-10-156-761-11850

Sequence 11850, Application US/10156761

Publication No. US20030119018A1

GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI

APPLICANT: IKEDA, HARUO

APPLICANT: ISHIKAWA, JUN

APPLICANT: HORIKAWA, HIROSHI

APPLICANT: SHIBA, TADAYOSHI

APPLICANT: SAKAKI, YOSHIYUKI

APPLICANT: HATTORI, MASAHITKA

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-262

CURRENT APPLICATION NUMBER: US/10/156,761

CURRENT FILING DATE: 2002-05-29

PRIOR APPLICATION NUMBER: JP 2001-204089

PRIOR FILING DATE: 2001-05-30

PRIOR APPLICATION NUMBER: JP 2001-272697

PRIOR FILING DATE: 2001-08-02

NUMBER OF SEQ ID NOS: 15109

SEQ ID NO 11850

LENGTH: 653

TYPE: PRF

ORGANISM: Streptomyces avermitilis

US-10-156-761-11850

Query Match

Best Local Similarity 39.2%; Score 100; DB 15; Length 653;

Matches 29; Conservative 15; Mismatches 20; Indels 10; Gaps 4;

QY 104 LHYGPGTGKSTFLGAINQLKSKVRSITITL--PEPRTLKGGFKGSGFEKRLHVR 157

Db 350 LFTYGGTGKTKTLHLAGVNGIMARKPKNVYMSERFQDVAKALQNNVAIEFFRRYYR 227

QY 158 HRYREANILMLDDI 171

Db 407 -RYREMDILLVDDI 419

RESULT 3

US-09-815-242-10395

Sequence 10395, Application US/09815242

Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Karl L.

APPLICANT: Zyekind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 10395

LENGTH: 467

TYPE: PRF

ORGANISM: Escherichia coli

US-09-815-242-10395

Query Match

Best Local Similarity 7.4%; Score 98; DB 9; Length 467;

Matches 19; Conservative 21; Mismatches 28; Indels 2; Gaps 1;

QY 104 LHYGPGTGKSTFLGAINQLKSKVRSITITL--PEPRTLKGGFKGSGFEKRLHVR 161

Db 168 LFTYGGTGKTKTLHLAGVNGIMARKPKNVYMSERFQDVAKALQNNVAIEFFRRYYR 227

QY 162 EANIIMLDDI 171

Db 228 SVDALILDDI 237

RESULT 4

US-09-815-242-13552

Sequence 13552, Application US/09815242

Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Karl L.

APPLICANT: Zyekind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 13552

LENGTH: 442

TYPE: PRF

ORGANISM: Streptococcus pneumoniae

US-09-815-242-13552

Query Match

Best Local Similarity 7.3%; Score 96.5; DB 9; Length 442;

Matches 42; Conservative 38; Mismatches 78; Indels 49; Gaps 7;

QY 50 ERFPAELITSHM---QRTLNALKK-----DIMNRHRRDLVAM 86

Db 56 EIVDAE-ITPHYITFTKQDTSSQVEATNLTLYDSPLVISPVDGKKEKTYFDNFI 114

QY 87 AADD-----ICTAITNGEYV-----KGYLYGPGTGKSTFLGAINQLKSKVRSITITL 138

Db 115 QGDGNWAVASALAVSEDLATFTNPLFIYQPGIGKTHLNLNIGNELKNIPARVYKIP 174  
Qy 139 --EPIRTLKGFGDGSFEKKLHVRREANILMDIGAEVTPWVRDEVIGPLHYRMYE 196  
Db 175 ASSEFIDFIDHRLGEMERFKTYRSLDLIDDIOS-----LSGKRVAT 219  
Qy 197 LPTFSSNFDYSELEHNLMTDRGEK 223  
Db 220 QEEFNTFMAIDKQKQIVLTSRSPK 246

## RESULT 5

US-09-815-242-11167  
; Sequence 11167, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FaastSeq for windows Version 4.0  
; SEQ ID NO 11167  
; LENGTH: 454  
; TYPE: PRY  
; ORGANISM: Haemophilus influenzae  
US-09-815-242-11167

Query Match 7.2%; Score 96; DB 9; Length 454;  
Best Local Similarity 24.0%; Pred. No. 0.095;  
Matches 46; Conservative 31; Mismatches 75; Indels 40; Gaps 9;

Qy 51 REAEELISHHQRDTLNKLDIYNNHDDRLDVMAADDICTAINGQVKGILYIGPF 110  
Db 114 KRESHLNTH-----LPDNFVEGKSN-QLAAVGQKLAQAGEPSANPFILYGET 162  
Qy 111 GTGKSEFIIIGAINOLKSKVRSIIYL--PEFIRTLKGFGKGSFE--KKLRVREANIL 166  
Db 163 GIGKTHLHAIIGNGIADKPNARVLYIHANNFMQHWKAVRNDKMDQFKFTRSLDA--L 220  
Qy 167 MLDI--GAEEVTPWVRDEVIGPLHYRMYHELPTFSSNFDYSELEHNLMTDRGEK 223  
Db 221 LVDDIQFPAKEKETQ-----EEFHIIFNSLF--ETGQIILTSRSPK 261  
Qy 224 TKAARIERVKS 235  
Db 262 -EIEKIERLKS 272

## RESULT 6

US-10-032-585-7201  
; Sequence 7201, Application US/10032585  
; Publication No. US20030180953A1  
; GENERAL INFORMATION:  
; APPLICANT: Terry, Roemer D.  
; APPLICANT: Bo, Jiang  
; APPLICANT: Charles, Boone  
; APPLICANT: Howard, Bussey  
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery  
; FILE REFERENCE: 10182-005-999  
; CURRENT APPLICATION NUMBER: US/10/032,585  
; CURRENT FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 8000  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 7201  
; LENGTH: 796  
; TYPE: PRY  
; ORGANISM: Candida albicans  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (383)..(383)  
; OTHER INFORMATION: X-any amino acid  
US-10-032-585-7201

Query Match 7.0%; Score 93.5; DB 12; Length 796;  
Best Local Similarity 26.5%; Pred. No. 0.41;  
Matches 40; Conservative 22; Mismatches 44; Indels 45; Gaps 8;

Qy 67 LNAKLDIY--MNR-----DRLDVMAADDICTAINGQVKGILYIGPFQKSP 116  
Db 282 LDAEFQDIFRAFNRIILPPELAEKLDY-----KHKGGILYGPFGTKS 327  
Qy 117 ILGAIANOLKSKVRSIIYLPEFIRTLKGC-----FKGSEKKLH-RVREANILM 167  
Db 328 IARKLSKMLNGKEPR--IVNGPEMLSKYVGASEENIRMLFDKAEKYLKGGDSLDLXII 385  
Qy 168 LDDI-----GAEEVTPWVRDEVIGPL 169  
Db 386 FDELDSVFKQSGSGSDGT-GVGDVAVNQL 415

## RESULT 7

US-09-764-868-774  
; Sequence 774, Application US/09764868  
; Patent No. US20020168711A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT232  
; CURRENT APPLICATION NUMBER: US/09/764,868  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PAM or file wrapper  
; NUMBER OF SEQ ID NOS: 1510  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 774  
; LENGTH: 574  
; TYPE: PRY  
; ORGANISM: Homo sapiens  
US-09-764-868-774

Query Match 7.0%; Score 92.5; DB 10; Length 574;  
Best Local Similarity 30.3%; Pred. No. 0.33;  
Matches 30; Conservative 17; Mismatches 33; Indels 19; Gaps 4;

Qy 102 KGLVYGFPGTKSFIIGAINOLKSKVRSIIYLPEFIRTLKGFGKGSFEKKLHRY- 160  
Db 127 RGVLLYGPFGTKMTAAVAVNEGA---YYSVINGPELISKFY-----GETEALRQIF 178  
Qy 161 -----REANILMDIGA-----EEVTPWVRDEVIGPL 189  
Db 179 AEATLRHPSIIFIDLDALCPREBGAONEVEKRVVASLL 217

## RESULT 8

US-10-032-585-7621  
Sequence 7621, Application US/10032585  
Publication No. US20030180953A1  
GENERAL INFORMATION:  
APPLICANT: Terry, Roemer D.  
APPLICANT: Bo, Jlang  
APPLICANT: Charles, Boone  
APPLICANT: Howard, Bussey  
TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery  
FILE REFERENCE: 10182-005-999  
CURRENT APPLICATION NUMBER: US/10/032,585  
NUMBER OF SEQ ID NOS: 8000  
SOFTWARE: PatentIn Version 3.1  
SEQ ID NO 7621  
LENGTH: 827  
TYPE: PRT  
ORGANISM: Candida albicans  
US-10-032-585-7621

## Query Match 6.8%; Score 91.5; DB 12; Length 827;

Best Local Similarity 31.8%; Pred. No. 0.71;  
Matches 35; Conservative 17; Mismatches 41; Indels 17; Gaps 6;

QY 96 TNGEYVGLYVPGTGTGKSFILGAIANOLKSKKVRSTIYLPFIRTLKGFDSGFEK 155  
DB 222 TNGEYVGLYVPGTGTGKSFILGAIANOLKSKKVRSTIYLPFIRTLKGFDSGFEK 155  
QY 156 KLRHV-REAN-----IMLDDIGAEBVTWVRDEVTGPLLHYRMVHELPT 199  
DB 274 KLRHFEAKQIAAPCLIFMDEIDA--ITP-KRDGAGQREMEKRIVAQLLT 320

## RESULT 9

US-09-815-242-10193  
Sequence 10193, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Hasebeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 10193  
LENGTH: 577  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-09-815-242-10193

Query Match 6.8%; Score 90.5; DB 9; Length 577;  
Best Local Similarity 18.0%; Pred. No. 0.55;  
Matches 68; Conservative 37; Mismatches 86; Indels 187; Gaps 14;

QY 1 YXDQQRHYD-GHKFAD-CENFYKGVHVELYVDNNRIRYLQCPCKIKYDEERFEALIT 58  
DB 193 YRDAGKGYDEDEFEARARVY-----VXLQSGDEYF----- 224  
QY 59 SHMQDQITNAKLKDYNNHR-----DRLDVMAADDICTAINTGSEYKGLYVGPRTGK 114  
DB 225 -REMR-----KLVDTYTONQITYDRLVNTLTRD-----VMBSLYNP----- 263  
QY 115 SFILGAIANOLKSKKVR-----STIYLPF-----IRTLKGF----- 148  
DB 264 --MLPGIVADVAKAGLAVERSEGATVVFDFEKNKEGEPGVIIQKXDDGGLYTTDIACA 321  
QY 149 -----KQGS----- 152  
DB 322 KYRYETLAHADVLYYIDSRQHLMQAWAIVRKAGYVPSVPLEHMFQMLGKDGKPFK 381  
QY 153 -----PEKTLHVRREANIMLDDIGAEBVTWVRDEVTG----- 186  
DB 382 TRAGTIVKLADLDELERARRLVAEKNDMPADELEKLANAVGICAVKADISKRTTD 441  
QY 187 -----PLHYRMVHELPTFFSSNPFYSELEHHLIANTRDGEETKARIT 230  
DB 442 YIFDWMNMLAFEGNTAPVQVATRYLSVFRKAEIDBQLAARPAVIRERERQALARRLL 501  
QY 231 ERYKLS-----TPYFL 242  
DB 502 QFETTLTVVAREBGTTPHVM 519

## RESULT 10

US-09-976-165-13  
Sequence 13, Application US/09976165  
Patent No. US20020107383A1  
GENERAL INFORMATION:  
APPLICANT: FUJIMURA, TSUTOMU  
APPLICANT: MATANABE, TAKESHI  
APPLICANT: HORIE, MASATO  
TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN  
FILE REFERENCE: Q-53599  
CURRENT APPLICATION NUMBER: US/09/976,165  
CURRENT FILING DATE: 2001-10-15  
PRIOR APPLICATION NUMBER: 09/565,538  
PRIOR FILING DATE: 2000-05-05  
PRIOR APPLICATION NUMBER: 09/055,699  
PRIOR FILING DATE: 1998-04-07  
PRIOR APPLICATION NUMBER: 08/820,170  
PRIOR FILING DATE: 1997-03-19  
PRIOR APPLICATION NUMBER: JP 63410/1996  
PRIOR FILING DATE: 1996-03-19  
PRIOR APPLICATION NUMBER: JP 69163/1997  
PRIOR FILING DATE: 1997-03-05  
NUMBER OF SEQ ID NOS: 95  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 13  
LENGTH: 389  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-976-165-13

Query Match 6.8%; Score 90; DB 10; Length 389;  
Best Local Similarity 35.1%; Pred. No. 0.35;  
Matches 27; Conservative 13; Mismatches 25; Indels 12; Gaps 3;

QY 102 KGLIYVPGTGTGKSFILGAIANOLKSKKVRSTIYLPFIRTLKGFDSGFEK 156  
DB 168 KGLIYVPGTGTGKSFILGAIANOLKSKKVRSTIYLPFIRTLKGFDSGFEK 156



```
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5010
LENGTH: 445
TYPE: PR1
ORGANISM: Enterococcus faecalis
US-09-815-242-5010
```

```
Query Match 6.8%; Score 90; DB 9; Length 445;
Best Local Similarity 22.1%; Pred. No. 0.42;
Matches 44; Conservative 30; Mismatches 55; Indels 70; Gaps 9;
```

```
QY 27 LTVNNRIKIRYQCPCKIKYD-----KDIYNNRDLVDVMAADDICTAITNG 98
DB 34 LKIDNNQ---WLEVPASVHRDYWEKNLSAKIVETGFKLTGAEMPHFVVADEKDALAQ 90
QY 49 --EEFPAELITSHHQRDTLNKL-----KDIYNNRDLVDVMAADDICTAITNG 98
DB 91 ELEPABEEVVFESQSKAMLNPKYTFDTPVIGKQNMHAAL---VVAED-----PG 141
QY 99 EQVKGILYGPFGTGSFIIAIAIQL---KSKYR--STIYLPFIRTLKGFGFGDS 152
DB 142 SIYNPLPFYGVGVGLKTHLMHAIGHQMLVNPQPAKVKYVSSFTFTFINSIQTKTSE-Q 200
QY 153 FEKKLRVREANILMLDDI 171
DB 201 FRKE---YRNVDDLVDI 216
```

```
RESULT 15
US-09-815-242-10908
Sequence 10908, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohleen, Karl L.
APPLICANT: Zyckind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
```

```
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10908
LENGTH: 447
TYPE: PR1
ORGANISM: Enterococcus faecalis
US-09-815-242-10908
```

```
Query Match 6.8%; Score 90; DB 9; Length 447;
Best Local Similarity 22.1%; Pred. No. 0.43;
Matches 44; Conservative 30; Mismatches 55; Indels 70; Gaps 9;
```

```
QY 27 LTVNNRIKIRYQCPCKIKYD-----KDIYNNRDLVDVMAADDICTAITNG 98
DB 34 LKIDNNQ---WLEVPASVHRDYWEKNLSAKIVETGFKLTGAEMPHFVVADEKDALAQ 90
QY 49 --EEFPAELITSHHQRDTLNKL-----KDIYNNRDLVDVMAADDICTAITNG 98
DB 91 ELEPABEEVVFESQSKAMLNPKYTFDTPVIGKQNMHAAL---VVAED-----PG 141
QY 99 EQVKGILYGPFGTGSFIIAIAIQL---KSKYR--STIYLPFIRTLKGFGFGDS 152
DB 142 SIYNPLPFYGVGVGLKTHLMHAIGHQMLVNPQPAKVKYVSSFTFTFINSIQTKTSE-Q 200
QY 153 FEKKLRVREANILMLDDI 171
DB 201 FRKE---YRNVDDLVDI 216
```

```
Search completed: December 15, 2003, 15:19:03
Job time : 25.4154 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 15:11:40 ; Search time 14.099 Seconds  
(without alignments)  
1705.235 Million cell updates/sec

Title: US-09-689-952-18

Perfect score: 1328

Sequence: 1 YDQOKHYDGHKFAQCPNFV.....ERVKSLSTPYFLSGENFRNN 250

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1328	100.0	306	2 A89952	primosomal protein
2	537.5	40.5	311	1 IQBS44	primosome componen
3	531.5	40.0	311	2 H84042	primosome componen
4	475.5	35.8	307	2 AB1632	primosome componen
5	472.5	35.6	307	2 AH1269	primosome componen
6	381	28.7	298	2 B98066	primosome componen
7	361	27.2	298	2 B95199	primosomal protein
8	342	25.8	293	2 H86718	primosomal protein
9	173	13.0	209	2 S77882	dnaa protein homol
10	145	10.9	282	2 D97138	DNA replication pr
11	140	10.5	316	2 B82907	conserved hypothe
12	138.5	10.4	235	2 E70378	DNA replication pr
13	127.5	9.6	313	2 D69945	phage-related proc
14	127	9.6	266	2 140411	PBSX prophage ORF
15	125.5	9.5	440	2 D81415	chromosomal replic
16	120.5	9.1	316	2 T10436	probable transposa
17	115.5	8.7	246	2 D84932	DNA replication pr
18	115	8.7	399	2 G70328	chromosome replica
19	114	8.6	265	1 BVEICIT	1stb protein - Bac
20	112	8.4	287	2 AD1588	similarities Staph
21	112	8.4	287	2 AG1651	similarities Staph
22	112	8.4	291	2 A86681	hypothetical prote
23	112	8.4	520	2 H69125	hypothetical prote
24	111.5	8.4	475	2 C96503	protein FecI6.7 li
25	111	8.4	350	2 B97735	probable ATPase n2
26	109.5	8.2	324	2 C86802	hypothetical prote
27	108.5	8.2	329	2 D97340	DNA replication pr
28	107	8.1	585	2 F90460	AAA family ATPase
29	106.5	8.0	251	2 S23889	hypothetical prote

30	106.5	8.0	630	2 AC1309	probable ABC trans
31	106	8.0	798	2 B71196	probable transactio
32	105.5	7.9	248	2 C85626	probable DNA repli
33	105.5	7.9	248	2 B90763	probable DNA repli
34	105.5	7.9	699	2 B72479	probable transactio
35	105	7.9	246	2 C90913	probable DNA repli
36	105	7.9	248	1 C64886	DNA replication pr
37	105	7.9	795	2 F75154	cell division cont
38	104.5	7.9	248	2 A85747	probable DNA repli
39	104.5	7.9	248	2 A99872	probable DNA repli
40	104.5	7.9	382	2 H72255	myo-inositol-1-pho
41	104.5	7.9	454	2 D84931	chromosomal replic
42	104	7.8	248	2 AC0738	probable DNA repli
43	104	7.8	454	1 IQJYBA	replication initia
44	103.5	7.8	264	2 T47074	hypothetical prote
45	103.5	7.8	726	2 D72613	probable transactio

#### ALIGNMENTS

RESULT 1  
A89952  
primosomal protein [imported] - Staphylococcus aureus (strain N315)  
C/Species: Staphylococcus aureus  
C/Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C/Accession: A89952  
R/Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi, A.; Mizutani-U, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kato, C.; Sekimizu, K.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratsugu, K.  
Lancet 357, 1225-1240, 2001  
A/Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A/Reference number: A89758; MUID:21311952; PMID:11418146  
A/Accession: A89952  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-306 <KUR>  
A/Cross-references: GB:BA000018; PID:g13701480; PIDN:BA842774.1; GSPDB:GN00149  
A/Experimental source: strain N315  
C/genetics:  
A/gene: dnaI  
A/superfamily: 44K dnaa protein homolog

Query Match 100.0%; Score 1328; DB 2; Length 306;  
Best Local Similarity 100.0%; Pred. No. 1.3e-100;  
Matches 250; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	YDQOKHYDGHKFAQCPNFVKGVPELVYDNNRIRIKRYLQCPCKIKYDERFEALITSH	60
DB	57	YDQOKHYDGHKFAQCPNFVKGVPELVYDNNRIRIKRYLQCPCKIKYDERFEALITSH	116
QY	61	HNQDPTLNAKLDIYNNHNRDLVMAADDICTTAITNGEQVGLYLPGFGTKSFLCA	120
DB	117	HNQDPTLNAKLDIYNNHNRDLVMAADDICTTAITNGEQVGLYLPGFGTKSFLCA	176
QY	121	INQKSKKVRSTIYLPFRITLKGFGDGSFEKKHVRANITIMDDIAEVTTPV	180
DB	177	INQKSKKVRSTIYLPFRITLKGFGDGSFEKKHVRANITIMDDIAEVTTPV	236
QY	181	RDEVIGPLHYMVHELPTFFSSNFDYSELHNLAMTRDGESEKTYAARIIEVKSLSTPY	240
DB	237	RDEVIGPLHYMVHELPTFFSSNFDYSELHNLAMTRDGESEKTYAARIIEVKSLSTPY	296
QY	241	FLSGENFRNN 250	
DB	297	FLSGENFRNN 306	

RESULT 2  
IQBS44  
primosome component (helicase loader) dnaI - Bacillus subtilis  
N/Alternate names: dnaa protein homolog, 44K; hypothetical protein Y (dnaB 3' region)  
C/Species: Bacillus subtilis

C/Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 19-Jan-2001  
C/Accession: B24720; C26580; F69617  
R/Organism: *B. subtilis*, N. Moriya, S. Mazza, P.G.; Yoshikawa, H.  
Nucleic Acids Res. 14, 9989-9999, 1986  
A/Title: Nucleotide sequence and organization of *dnab* gene and neighbouring genes on the  
A/Reference number: A93650; MUID:87117549; PMID:3027671  
A/Accession: B24720  
A/Molecule type: DNA  
A/Residues: 1331 <GGA>  
A/Cross-references: GB:X04963; NID:939880; PIDN:CAA28633.1; PID:G39881  
R/Hoshino, T.; McKenzie, T.; Schmidt, S.; Tanaka, T.; Sueoka, N.  
Proc. Natl. Acad. Sci. U.S.A. 84, 653-657, 1987  
A/Title: Nucleotide sequence of *Bacillus subtilis* *dnab*: a gene essential for DNA replication  
A/Reference number: A94709; MUID:87118226; PMID:3027697  
A/Accession: C26580  
A/Molecule type: DNA  
A/Residues: 1-18, 'N', 20-23, 'T', 25-206 <HOS>  
R/Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Beret  
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd  
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A/Authors: Foulger, D.; Fultz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gallier  
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holleppel, S.; Hosono, S.; Hullo, M.F.  
Koeter, P.; Koningsreth, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
A.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel  
Y. M.; Ogawa, K.; Ogilwa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelid  
Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanton,  
A.; Authors: Schleicher, S.; Schreier, R.; Scifone, F.; Sekiguchi, J.; Sekowska, A.; Serd  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpestra, P.; Tognoni, A.; Toesato, V.; Uchiyama,  
T.; Walters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
A/Authors: Yoshikawa, H.F.; Zumsen, E.; Yoshikawa, H.; Danchin, A.  
A/Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A/Reference number: A69580; MUID:98044033; PMID:9384377  
A/Accession: F69617  
A/Status: nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-311 <KUN>  
A/Cross-references: GB:Z99118; GB:AL009126; NID:92635200; PIDN:CA14856.1; PID:92635363  
A/Experimental source: strain 168  
C/Genetics:  
A/Gene: *dnai*  
C/Superfamily: 44k *dnai* protein homolog  
C/Keywords: ATP; nucleotide binding; P-loop  
F/168-175/Region: nucleotide-binding motif A (P-loop)  
Query Match 40.5%; Score 537.5; DB 1; Length 311;  
Best Local Similarity 44.9%; Pred. No. 2.8e-36;  
Matches 106; Conservative 48; Mismatches 81; Indels 1; Gaps 1;  
QY 15 DCPNPKVHPELVYNNRIRIKIRYLOCPCKIKYDEERFEALITSHHMDTLNAKLDI 74  
DB 75 NCNNLEIGHHPKLVNNGRSIDIEYECVPKRDQCKQKQSLKMSMTIQDLDGATQQV 134  
QY 75 YNNHRDLVAMADICTAITNGEYKGLYLPFGTGKSPILGAIANOLSKKVRSTI 134  
DB 135 DISDPERLAFQVHTVPLKSYNETGKGLYLVGKGVGTFLMALANLAKKSVSWI 194  
QY 135 IYPERITLKGFGKGSFFKKLHRYREANITLMDIGAEVTPWVRDEVYIGPLHYRMV 194  
DB 195 VYVPEFVRELKNSLOQDTLEKLNMYKTPVLMLDIGAESMTSWRDEVYIGVTLQHRMS 254  
QY 195 HELPTFFSSNFVDSLEHNLAMTRDGE-EKTKAARIERKSLSTPFTSGENFRN 249  
DB 255 QQLPTFFSSNFSPDELGHHTTYSORGEKEVKARLMERILYLAAPRIIDGENRRH 310

## RESULT 3

H84042  
primosome component (helicase loader) *dnai* [imported] - *Bacillus halodurans* (strain C-12)  
C/Species: *Bacillus halodurans*  
C/Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C/Accession: H84042  
R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai

Nucleic Acids Res. 28, 4317-4331, 2000  
A/Title: Complete genome sequence of the alkaliophilic bacterium *Bacillus halodurans* and  
A/Reference number: A83650; MUID:20512582; PMID:11058132  
A/Accession: H84042  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-311 <STO>  
A/Cross-references: GB:AP001517; GB:BA000004; NID:910175500; PIDN:BA006863.1; GSPDB:GN001  
A/Experimental source: strain C-125  
C/Genetics:  
A/Gene: *dnai*  
C/Superfamily: 44k *dnai* protein homolog

Query Match 40.0%; Score 531.5; DB 2; Length 311;  
Best Local Similarity 41.2%; Pred. No. 8.5e-36;  
Matches 107; Conservative 55; Mismatches 79; Indels 19; Gaps 6;  
QY 1 YKDOQKHVDGH--KPADCPNFVKGHPYELVYNNRIRIKIRYLOCPCKIKYDEERFEALIT 58  
DB 59 YKQKQSHC-AHCQGLQKCNLMKGYPTLYVERDSLELSYSPDLKEERREKKRSLR 117  
QY 59 SHHMDTLNAKLDI-----YNNHRDLVAMADICTAITNGEYKGLYLPFG 111  
DB 118 SLVTPKEILAEKFDVSESPGRSIAHR-ALBFLSAKP-----GEGMGILYKQDG 169  
QY 112 TGKSPILGAIANOLSKKVRSTIYPERITLKGFGKSPFKLHRYREANITLMDI 171  
DB 170 VGTFTPMGAILNELKRGIDSTIVYDPDFRELKQSGIDGTQKLDPVNMQVLLFDPI 229  
QY 172 GAEVTPWVRDEVYIGPLHYRWVHELPTFFSSNFDSLEHNLAMT-RDGEKTKAARI 230  
DB 230 GAEVTPWVRDEVYIGPLHYRWVHELPTFFSSNFDSLEHNLAMT-RDGEKTKAARI 289

QY 231 ERYKSLSTPFTSGENFRN 250  
DB 290 ERIHRYTVSVWVGQNRREH 309

## RESULT 4

AB1632  
primosome component (helicase loader) *dnai* [imported] - *Listeria innocua* (strain C11p112)  
C/Species: *Listeria innocua*  
C/Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C/Accession: AB1632  
R/Glaeser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duesurget, O.; Entian, K.D.; Feiht, H.;  
Science 294, 849-852, 2001  
A/Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maltournam, A.; Mat  
ok, C.; Schlueder, T.; Simoes, N.; Tillerer, A.; Vazquez-Boland, J.A.; Voss, H.; Weiland,  
A/Title: Comparative genomes of *Listeria* species  
A/Reference number: AB1077; MUID:21537279; PMID:11679669  
A/Accession: AB1632  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-307 <GLA>  
A/Cross-references: GB:AL592022; PIDN:CMC96826.1; PID:g16414082; GSPDB:GN00178  
A/Experimental source: strain C11p11262  
C/Genetics:  
A/Gene: *dnai*  
C/Superfamily: 44k *dnai* protein homolog

Query Match 35.8%; Score 475.5; DB 2; Length 307;  
Best Local Similarity 42.6%; Pred. No. 3e-31;  
Matches 98; Conservative 45; Mismatches 84; Indels 3; Gaps 2;  
QY 22 GHVPELVYNNRIRIKIRYLOCPCKIKYDEERFEALITSHHMDTLNAKLDIYNNHRDR 81  
DB 77 GYAPKLVNGBFITYYITKKEIEDKRAVERRRISRYMKQVVDNALADPFTYDEBR 136

QY 82 LDVAMADICTAIT--TNGEYKGLYLPFGTGKSPILGAIANOLSKKVRSTIYLP 139  
DB 137 KVALVAYQFLNNYPPQNERVKGFLTHSFGKSKYLLGALAKELALNGISTVLYLPE 196

Qy	Db	Qy	Db
140	FIRLKGFGKSGSEKTLHVR	140	FIRLKGFGKSGSEKTLHVR
141	ENIMLMDIGAEVLPWNR	141	ENIMLMDIGAEVLPWNR
142	SVIGLILHYRVHLEPT	142	SVIGLILHYRVHLEPT
143	197	197	197
197	FMRVKQSIDNVTAGEKIFAK	197	FMRVKQSIDNVTAGEKIFAK
198	EVLMDDIGAESMTAWR	198	EVLMDDIGAESMTAWR
199	DEVLGIQFRQGEELPT	199	DEVLGIQFRQGEELPT
200	200	200	200
201	PFSSNPDISELHILATNRG	201	PFSSNPDISELHILATNRG
202	-EERTKQARIERYKSLSTY	202	-EERTKQARIERYKSLSTY
203	PFSGENR	203	PFSGENR
204	257	257	257
257	PFSSNNMDLENTLMFAONG	257	PFSSNNMDLENTLMFAONG
258	STEELKARRIMEEVRYSK	258	STEELKARRIMEEVRYSK
259	SVNLEGNNR	259	SVNLEGNNR
306	306	306	306

## RESULT 5

primosome component (helicase loader) DnaI [imported] - *Listeria monocytogenes* (strain  
C:Species: *Listeria monocytogenes*  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C:Accession: AH1269  
R:Glaser, F.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entliam, K.D.; Faith, H.H.;  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkai, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schluter, T.; Simoes, N.; Tarrere, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
A:Title: Comparative genomics of *Listeria* species  
A:Reference number: AB1077, MUID:21537279, PMID:11679669  
A:Accession: AH1269  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-307 #Glas  
A:Cross-references: GB:NC 003210; PIDN:CAIC9638.1; PID:g1410989; GSPDB:GN0177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: dnaI  
C:Superfamily: 44k dnaI protein homolog

Query Match	35.6%	Score 472.5	DB 2	Length 307
Best Local	42.1%	Pred.No. 5.3e-31		
Matches 98	Conservative 47	Mismatches 79	Indels 9	Gaps 3

Qy	22 GHVBELYVDNNRIKIRYLÖCPCKIKYDEERPEALLTSHNÖRDTLNAKLKOIYMNHDR	81
	:   :   :   :   :   :   :   :   :   :   :   :   :	
Db	77 GYAKLVNLNEEFITVTYYPTXKEIEEDKRRAVERRTLSLMYKQVDANLADFTYDSESR	136

```
QY      82 LDVMAADICTAINTN-----GEQVGKLYLXGPFPGSKFILGAIANOLKSKVRSTIY 136
```

Qy 137 LPBFIRTLKGFKDSFEKLLHRVREANILMLDDIGAEVTPWVRDEYIGLLHYRWHE 196

197 LPTPFSSNFDYSELEHHLAMTRDG-EETKAARTIERVKSLSTPYFLSGENFR 248

D0 254 LPTFFSSNTYMDQLENNHLMFAQNGTEBKLKARRIMERVRYLSKEVNI EGKNRR 306

## RESULT 6

primosome component (helicase loader) [imported] - *Streptococcus pneumoniae* (strain R6)  
C/Species: *Streptococcus pneumoniae*  
C/Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 02-Nov-2001  
C/Accession: 1989066  
R/Hoskins, J.A., Alborn Jr., W., Arnold, J., Blaszcak, L., Burgitt, S., DeHoff, B.S., E  
y, R., LeBlanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P., McAhren, S., M  
y, P., Sun, P.M., Winkler, M.E., 2001  
J. Bacteriol. 183, 5709-5717, 2001  
A/Authors: Yang, I., Young-Bellido, M., Zhao, G., Zook, C., Baltz, R.H., Jaskunas, S.R.,  
A/Rtitle: Genome of the Bacterium *Streptococcus pneumoniae* Strain R6.  
A/Reference number: A97872, MUID:21432245, PMID:11544234

A, Status: preliminary  
A, Molecule type: DNA  
A, Residues: 1-298 <KUR>  
A, Cross-references: GB:AE007317, PIDN:AAU00359.1, PID:G154559220, GSPDB:GN00174

C;Genetics: A;Gene: dnaI C;Superfamily: 44K dnaA protein homolog

Query Match 28.7%; Score 381; DB 2; Length 288;  
Best Local Similarity 34.7%; Pred. No. 1,4e-23;  
Matches 82; Conservative 50; Mismatches 100; Indels 4; Gaps 3

15 DCPNFVKGHVPELVYDNNRIKIRYLCQCPCKIKYDEERFEALITSHHMQRDTLNAKLDI 74

Db 65 DTDYAKGYKPILVNHHGYADVSEETPELTAEKKAIAKKRLNLINPSSLKVNSFLDV 12

75 YMHNRDLVMAADDICTAITNGEQVGLYLYGPFQGRKSLIGALANQLKSK-VRST 13

DB 125 YRDDVQRLTVLAKRMIEFVNDYFN--NLKGLYLGDFGVGKSFVVAALAHDLSEKRGVSSI 18

[illegible]

104 רחוקה ונפרדת זכות יצירה/הערה: 104

||||| : | | : | | : | | :  
||| : || : || : || : || :  
242 OENT DTBETSNENFEOT FKJPA KIKHGNDEMTWABRIMEDTBYV A EFTBI FGJNBPB 298

## RESULT 7

primosomal protein DnaI [imported] - Streptococcus pneumoniae (strain TIGR4)  
C1species: Streptococcus pneumoniae  
C1date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 24-Aug-2001

R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heidelberg, J.D.; Imamura, T.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, R.;

erson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001

A;Authors: Iottus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison, A.;Title: Complete Genome Sequence of a virulent isolate of *Streptococcus pneumoniae*.

A/Accession: D95199  
A/Reference number: A55000; MUID:2135/203; PMID:11403918

```

A;Molecule type: DNA

```

A; Cross-references: GB:AE005672; P. A: Experimental source: strain TIGP

C;Genetics:  
A:Gene. SP1711

C;Superfamily: 44K dnaA protein homolog

Query Match	21.2%;	Score 361;	DB 2;	Length 298;
Best Local Similarity	34.7%;	Pred. No. 6e-22;		

malicious	50%	malicious	50%	diverse	10%	other	3%
conservative	50%	malicious	50%	diverse	10%	other	3%

[illegible]

75 YANIBDPI DV--ZAMBADICTATINCEOVKCI YI VCEBETCKSEITIGAITANOI KSKK-V 130

Db 125 DLDLGRLPVEKLLAFVBOYPAIR-----KGLYLQDPEGVGKSFVWALLAHLSEKRV 179

131 RSTIIYPFIRTLKGGFKDGSFEKCLHRVREANILMLDDIGAEVTPWVRDEVIGPLH 190

Db 180 S<sup>1</sup>TLHYP<sup>2</sup>SFVIDVKA<sup>3</sup>ISDGN<sup>4</sup>VT<sup>5</sup>LVDEIK<sup>6</sup>SEVL<sup>7</sup>ILDDI<sup>8</sup>GAEQ<sup>9</sup>STVWVR<sup>10</sup>DEILQ<sup>11</sup>VILQ<sup>12</sup> 239

QY 191 YRMVHELPTFFSSNFDYSEIHLAMTRDGEKT-KAARTERVKSLSTPYFLSGENFR 248

Db 240 YRQGENLPFTSFNFEDLEKHPAKVKHGNDETWEARRVMERIRYLAETRLEGVNR 298

## RESULT 8

primosomal protein DnaI (imported) - *Lactococcus lactis* subsp. *lactis* (strain IL1403)  
C1:Species: *Lactococcus lactis* subsp. *lactis*



Query Match 31.0.4%; Score 138.5; DB 2; Length 235;  
Best Local Similarity 23.8%; Pred. No.0.00056;  
Matches 59; Conservative 41; Mismatches 93; Indels 55; Gaps 9;

```
QY      19 FVKGHVEPLVYDNNRIKIRYIOCPCKIKYDERPEAEILITSHMORTLNAKLDYIMNH 78
Dd      15 FVKK-----FEDNKVRL-----CECRFPKRRVNNREL-----NIPKRYNNANL-DYHPK 56
QY      79 RDRLDVMAAADDCIATNGEIOVKGLYLPGEPGKSPITGAINLNOLSKK-VRSITIIYL 137
Dd      57 NVSQNRALLITIRVPHANNPPEKGLTFVSGPGCKHTLAATLKAIVKKGIGHGYPFD 116
QY      138 PEPIRTLKGFRDGSFEKCLHRVREANILMLDIGAEEFTWVDEVIQPLLHYRMVEL 197
Dd      117 KDLFRLGLHDEGDKOTFKTLTVLNSPVLVYDDIGSERLSMWOR-ELISYITIRYNNLK 175
QY      198 PTFSSNDVSELEHILAMTRDGEKTK-----AARIERY----- 233
Dd      176 STITTTT-----SLQREBSVSRIADLARSLEGNVSKIYENNELVIKSGDLR 226
QY      234 --KSLSTP 239
Dd      227 KSKKLSTP 234
```

RESULT 13

D69945  
phage-related protein homolog yqam - Bacillus subtilis  
C1Species: Bacillus subtilis  
C1Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999  
C1Accession: D69945

A1Authors: Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertele, C., Bron, S., Brulliet, S., Brusch, C.V., Caldwell, B., Capuano, V., Carter, N.M., Chai, A., Ehrlich, S.D., Emmerzon, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrari, E., Nature 390, 249-256, 1997

A2Authors: Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Funo, S., Galizzi, A., Gallier, J., Harwood, C.R., Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F., Jach, J., Kohnigstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Lardinois, Koester, P., Koningsfeld, J., Lazarovic, V., Lee, S.M., Levine, A., Liu, H., Masuda, A1Authors: Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H., Masuda, S., Maunel, Y. M., Ogawa, K., Ogiwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M., Portecelle, Rieger, M., Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadaie, Y., Sato, T., Scanlon, A1Authors: Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J., Sekowska, A., Serotoni, E.

```

Query Match 9.6%; Score 127.5; DB 2; Length 313;
Best Local Similarity 33.1%; Pred. No. 0.0063;
Matches 40; Conservative 19; Mismatches 49; Indels 13; Gaps 6;

QY 98 GEQVKGALVYGPFGTKSFILGAIANQL-KSKRVRSITLYLP--EFIRTLKGGFKDGSFE 154
DB 161 GERQSIALLGQPGSGKTHLTALIMNNLLIKKSSVH--CWYFPVVEGGDGLKAPF--DNLE 216

QY 155 KKLHVRVRANILIMDDI-----GAEVTPWMDDEVIGPLHYHNMHELPFFSGSNPFYSE 209
DB 217 AKLDMARKVEVLFIDLLFKRPINGQPRATDW-QVEQIQSVLNYRYLNHKKPLISSSELTIDE 275

QY 210 L 210
DB 276 I 276

```

RESULT 14  
140411  
PBsx prophage ORF xkdc - Bacillus subtilis  
N|Alternate names: protein S (xre region)  
C|Species: Bacillus subtilis  
C|Date: 12-Aug-1996 #sequence revision 12-Aug-1996 #text change 15-Oct-1999  
C|Accession: 140411; C69731; G47111  
R.|McDonnell, G.E.; Wood, H.; Devine, K.M.; McDonnell, D.J.  
J. Bacteriol. 176, 5820-5830, 1994  
A|Title: Genetic control of bacterial suicide: regulation of the induction of PBsx in Bacillus subtilis  
A|Reference number: 140408; MUID:94364963; PMID:8083174  
A|Accession: 140411  
A|Status: preliminary; translated from GB/EMBL/DDBJ  
A|Molecule type: DNA  
A|Residues: 1-266 <RES>  
A|Cross-references: EMBL:Z34287; NID:g498810; PIDN:CAB84044.1; PID:g498814  
R.|Kunst, F.; Ogasaawa, N.; Mosser, I.; Albertini, A.M.; Allioni, G.; Avevedo, V.; Berteroglio, V.; Bruch, J.; Bron, S.; Boullet, S.; Brusch, C.V.; Caldwell, B.; Capunov, V.; Carer, N.M.; Choi, C.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A|Authors: Fouger, D.; Filiz, C.; Fujita, M.; Fujita, Y.; Puma, S.; Gallizi, A.; Gallero, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Husono, S.; Hullio, M.F.; Koetter, J.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidos, A.; Lardinois, J.A.; Lauber, J.; Lazarovic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mausel, Y.M.; Ogawa, K.; Ogiwara, A.; Ondaea, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Saaipe, Y.; Sato, I.; Scanlon, A.; Authors: Schlecht, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Setor, T.; Winters, P.; Wipac, A.; Tanaka, T.; Tepstra, P.; Togomi, A.; Tosato, V.; Uchiyama, A.; Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A|Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A|Reference number: A65980; MUID:98044033; PMID:9384377  
A|Accession: C69731  
A|Status: nucleic acid sequence not shown; translation not shown  
A|Molecule type: DNA  
A|Residues: 1-238, 'K', 240-266 <KUN>  
A|Cross-references: GB:Z99110; GB:ALU09126; NID:g2633472; PIDN:CAB13110.1; PID:e1183273;  
C|Genomic source: strain 168  
C|Genetics:  
A|Gene: xkdc

Query Match                  9.6%; Score 127; DB 2; Length 266;

Best Local Similarity 31.0%; Pred. No. 0.0056;  
Matches 39; Conservative 25; Mismatches 44; Indels 18; Gaps 7;

```
QY          CTA--ITNGEYK-----GLVYFGPGTGSFTLGIAMQLSKRYSTIYP--EFIR 142
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db          CTKEVADAYEQIKDCKRKSIALLGGPSSGSKTHLLTAANEL-MRTCYVPVIYFFPEEGF 159
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY          TLKGGFKDGSPEKKLHVRREANIIMLDI----GAEEVTPWVRDEVIGPLHYRWVHEI 197
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db          DLKNDF--ALLERAKLNRMKQADVFLIDDLFKPVNKRPRATDWOLEMYS-VLNTYRIYNHK 216
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY          LPSGPTFFSS 203
           |||
Db          FILISS 222
```

RESULT 15

chromosomal replication initiator protein Cj0001 [Imported] - Campylobacter jejuni (strain  
C/Species: Campylobacter jejuni)  
C/Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
C/Accession: D81415  
R./Parkhill, J.; Wren, B.W., Mungall, K., Ketley, J.M., Churcher, C., Basham, D.; Chillingworth,  
C.W.; Quail, M.; Rajadurai, M.A.; Kutherford, R.M.; VanVlier, A.; Whitehead, S.; Barrett,  
Nature 403, 665-668, 2000  
A./Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyv  
A./Reference number: AB1250; MUID:20150912; PMID:10688204  
A./Accession: D81415  
A./Status: Preliminary  
A./Molecule type: DNA  
A./Residues: 1-440.<PAR>  
A./Cross-references: GB:AL139074; GB:AL111168; NID:66967505; PIND:CBM72494.1; PID:g6656750  
A./Experimental source: serotype O2, strain NCTC 11168  
C./Genetics:  
A./Gene: dnaA, Cj0001  
A./Superfamily: replication initiation protein dnaA

Query Match	9.5%;	Score 125.5;	DB 2;	Length 440;
Best Local Similarity	26.8%;	Pred. No. 0.014;		
Matches	41;	Conservative	30;	Mismatches 51;
				Indels 31;
				Gaps 7;

```

Qy 92 CTAITNGEONKGLY-----LYGPGGKSPFIIAGINQKSKVSTIYL--PEPIITLK 145
Db 123 CKRIAHKDKIGKLYNPIFYVGPPTGKTHLLQAVGN--ASLEMGKVIYATSENFINDP 180
Qy 146 GGPFGDSPEFKLHRVEANIILMDIGAEVLPWPRDVIQPLLHMYMHELPPFSNPF 205
Db 181 SNKNSGSDKFEHKRNCVLIIDVQVGLGTIDKIGEE-----FF---P 221
Qy 206 DYSELEHH--LAMPDGEKTKAARIYERYS 235
Db 222 IFNEIKNNDGQIIIMTSDNPNNLKG-ITERLKS 253

```

Search completed: December 15, 2003, 15:16:50  
Job time : 15.099 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 15:11:40 ; Search time 7.90922 Seconds

(without alignments)  
1486.453 Million cell updates/sec

Title: US-09-689-952-18  
Perfect score: 1328  
Sequence: 1 YKDQKHYDHGKFKADCPNFV.....BRVKSLSYFPLSGENFRNN 250

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues  
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41.\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	537.5	40.5	311	1	DNAI_BACSU
2	128.5	9.7	457	1	DNAI_BUCBP
3	127.5	9.6	313	1	YOAM_BACSU
4	127	9.6	266	1	XKOC_BACSU
5	123.5	9.5	440	1	DNAI_CAMUE
6	122	9.2	246	1	DNAI_BUCAP
7	115.5	8.7	246	1	DNAI_BUCAP
8	115	8.7	399	1	DNAI_AQUAE
9	114	8.6	265	1	ISTB_PSEAE
10	111.5	8.4	445	1	DNAI_ACHLA
11	109.5	8.2	229	1	ISTB_BURCE
12	107	8.1	451	1	DNAI_PASMU
13	106.5	8.0	251	1	ISTB_PASMU
14	105.5	7.9	231	1	ISTB_PSEFL
15	105	7.9	248	1	YDAP_ECOLI
16	104.5	7.9	454	1	DNAI_BUCAI
17	104	7.8	454	1	DNAI_BUCAP
18	103.5	7.8	263	1	Y4BW_RHISN
19	102.5	7.7	432	1	SKDI_SCHPO
20	102.5	7.7	466	1	DNAI_PROMI
21	102	7.7	462	1	DNAI_YERPE
22	102	7.7	464	1	DNAI_SERMA
23	100	7.5	263	1	ISTB_BACFR
24	100	7.5	307	1	RUVA_MYCGE
25	100	7.5	643	1	DNAI_STRRE
26	100	7.5	656	1	DNAI_STRCO
27	99.5	7.5	794	1	SC18_CANAL
28	99	7.5	4594	1	DYHC_HUMAN
29	99	7.5	4644	1	DYHC_MOUSE
30	99	7.5	4644	1	DYHC_RAT
31	98.5	7.4	453	1	DNAI_STRPN
32	98	7.4	245	1	DNAI_ECOLI
33	98	7.4	466	1	DNAI_SALTI

34	98	7.4	466	1	DNAI_SALTY	P35891 salmonella
35	98	7.4	467	1	DNAI_ECO57	O8xb23 escherichia
36	98	7.4	467	1	DNAI_ECOLI	P03004 escherichia
37	98	7.4	468	1	DNAI_VIBPA	O87eq7 vibrio para
38	97.5	7.3	449	1	DNAI_BACHD	O9rc02 bacillus ha
39	97.5	7.3	468	1	DNAI_VIBVU	O8dd19 vibrio vuln
40	97	7.3	340	1	RFC5_HUMAN	P40937 homo sapien
41	97	7.3	518	1	DNAI_NEIMB	O9jks7 neisseria m
42	97	7.3	2153	1	POIG_HRV16	O82122 human rhin
43	96.5	7.3	440	1	PR54_CHICK	O90732 gallus gall
44	96	7.2	410	1	PSMR_METTH	O26824 methanobact
45	96	7.2	436	1	PSMR_METTXA	O8rx03 methanopyru

## ALIGNMENTS

RESULT 1	ID	DNAI_BACSU	STANDARD;	PRT;	311 AA.
AC	P06567;				
DT	01-JAN-1988 (rel. 06, Created)				
DT	01-JAN-1988 (rel. 06, Last sequence update)				
DT	28-FEB-2003 (rel. 41, Last annotation update)				
DE	Primosomal protein dnaI.				
GN	DNAI				
OS	Bacillus subtilis.				
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.				
OX	NCBI_Taxid=1423;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=87117549; PubMed=3027671;				
RA	Ogasawara N., Moriya S., Maza P.G., Yoshikawa H.;				
RT	"Nucleotide sequence and organization of dnaB gene and neighbouring				
RT	genes on the Bacillus subtilis chromosome."				
RL	Nucleic Acids Res. 14:9989-9999(1986).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	STRAIN=168;				
RX	MEDLINE=97124191; PubMed=8969504;				
RA	Wipat A., Carter N., Brignell C.S., Guy J.B., Piper K.,				
RA	Sanders J., Emerson P.T., Harwood C.R.;				
RT	"The dnaB-rhea (256 degrees-240 degrees) region of the Bacillus				
RT	subtilis chromosome containing genes responsible for stress				
RT	responses, the utilization of plant cell walls and primary				
RT	metabolism."				
RL	Microbiology 142:3067-3078(1996).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=168;				
RX	MEDLINE=98048467; PubMed=9387221;				
RA	Lapidis A., Galleron N., Sorokin A., Ehrlich S.D.;				
RT	"Sequencing and functional annotation of the Bacillus subtilis genes				
RT	in the 200 kb rmb-dnaB region."				
RL	Microbiology 143:3431-3441(1997).				
RN	[4]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=168;				
RX	MEDLINE=98044033; PubMed=9384377;				
RA	Kuntz F., Ogasawara N., Mosser I., Albertini A.M., Alloni G.,				
RA	Azeredo V., Barreiro M.G., Bessieres C., Bolotin A., Borchert S.,				
RA	Borriero R., Bourbier L., Brans A., Braun M., Brignell S.C., Bron S.,				
RA	Brouillet S., Brusch C.V., Caldwell I.F., Cummins N.J., Carter N.M.,				
RA	Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,				
RA	Deutzot F., Devigne K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,				
RA	Enlison K.D., Errington J., Fabret C., Ferrari B., Foulger D.,				
RA	Fritz C., Fujita M., Fujita Y., Fuma S., Gallizon A., Galleron N.,				
RA	Ghim S.Y., Glaser P., Goffeau A., Colighly E.J., Grandi G.,				
RA	Ginepropi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,				
RA	Hilbert H., Holsappel S., Hosono S., Hullo M.F., Ilaya M., Jones L.,				
RA	Joris B., Karamata D., Kasahara Y., Kjaer-Blanchard M., Klein C.,				
RA	Kobayashi Y., Koester P., Konigstein G., Krogh S., Kumano M.,				
RA	Kurita K., Lapidis A., Lardinois S., Lauber J., Lazarevic V.,				

RA Lee S.M., Levine A., Liu H., Maeda S., Mauel C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
RA Nodine D., O'Reilly M., Ogawa A., Ogiwara S., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portetelle D., Portolillo S., Prescott A.M.,  
RA Pressac E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Roche B., Roche M., Rose M., Sadleir Y.,  
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
RA Sorochin A., Taccanti B., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpetre P., Tognoni A.,  
RA Tosoato V., Uchiyama S., Vandenbol M., Vamler P., Vassartolci A.,  
RA Viali A., Wambutt R., Wedler B., Wedler H., Weitzengesser T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,  
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
RT *subtilis* ";  
RL Nature 390:249-256(1997).  
RN [5]  
RP SEQUENCE OF 1-206 FROM N.A.  
RX MEDLINE=67118226; PubMed=3027697;  
RA Hoshino T., McKenzie T., Schmidt S., Tanaka T., Sueoka N.,  
RT "Nucleotide sequence of *Bacillus subtilis* dnaB: a gene essential for  
RT DNA replication initiation and membrane attachment";  
RL Proc. Natl. Acad. Sci. U.S.A. 84:653-657(1987).  
RN [6]  
RP SIMILARITY TO DNA.  
RX MEDLINE=92195821; PubMed=1549481;  
RA Koonin E.V., "Archaeobacterial virus SSV1 encodes a putative DnaA-like protein";  
RT Nucleic Acids Res. 20:1143-1143(1992).  
RN [7]  
RP IDENTIFICATION.  
RX MEDLINE=95291463; PubMed=7773414;  
RA Bruand C., Ehrlich S.D.,  
RT "The *Bacillus subtilis* dnaI gene is part of the dnaB operon";  
RL Microbiology 141:1199-1200(1995).  
CC -1- FUNCTION: PROBABLY INVOLVED IN DNA REPLICATION.  
CC -1- SIMILARITY: SOME, TO DNA FROM VARIOUS BACTERIA.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X04963; CA28633.1; -;  
DR EMBL; 275208; CA99605.1; -;  
DR EMBL; AF008220; AAC00359.1; -;  
DR EMBL; Z99118; CAB14858.1; -;  
DR EMBL; M15183; AAA22405.1; -;  
DR EMBL; B24720; I08544.  
DR Subtilin; BG10359; dnaI.  
DR InterPro; IPR003593; AAA\_ATPase.  
DR SMART; SMO0382; AAA; 1.  
KW Primosome; DNA replication; ATP-binding; Complete proteome.  
FT NP\_BIND 168 175 ATP (PROBABLY).  
FT CONFLICT 19 19 K -> N (IN REF. 5).  
FT CONFLICT 24 24 M -> T (IN REF. 5).  
SQ SEQUENCE 311 AA; 36114 MW; A86FC94AB6841264 CRC64;  
Query Match 40.5%; Score 537.5; DB 1; Length 311;  
Best Local Similarity 44.9%; Pred. No. 3.3e-37;  
Matches 106; Conservative 46; Mismatches 81; Indels 1; Gaps 1;

QY 135 IYHPEFRTLKGGFKDGSPEKTLHVRANILMLDDIGAEVTPWYRDEVIGPLHYRMV 194  
DB 195 VYVEFVRRLKSLQQTLEKKNMVKTPVLMDDIGESMTSWRDEVIGVLRHMS 254  
QY 195 HELPTFSSNFDYSBLEHMLAMTRDE-EKTKAARIERYKSISTPYPLSGENFRN 249  
DB 255 QQLPTFFSSNFDLSLKHFTYQSRQEKKEVKARILMERILYLAAPRLDGERRH 310  
RESULT 2  
ID DNA\_BUCBP STANDARD; PRT; 457 AA.  
AC P59567;  
DT 15-SRP-2003 (Rel. 42, Created)  
DT 15-SRP-2003 (Rel. 42, Last sequence update)  
DE 15-SRP-2003 (Rel. 42, Last annotation update)  
DE Chromosomal replication initiator protein dnaA.  
GN DNA OR BBP012.  
OS Buchnera aphidicola (subsp. *Batzongia pisiciferae*).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Buchnera.  
OX NCBI\_TaxID=135842;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22426901; PubMed=12522265;  
RA Van Ham R.C.H.J., Kamerbeek J., Palacios C., Rausell C., Abascal F.,  
RA Bastolla U., Fernandez J.M., Jimenez L., Postigo M., Silva P.J.,  
RA Tamames J., Viguera E., Latorre A., Valencia A., Moran F., Moya A.,  
RT "Reductive genome evolution in *Buchnera aphidicola*.";  
RL Proc. Natl. Acad. Sci. U.S.A. 100:581-586(2003).  
CC -1- FUNCTION: Plays an important role in the initiation and regulation  
CC of chromosomal replication. Binds to the origin of replication; it  
CC binds specifically double-stranded DNA at a 9 bp consensus (dnaA  
CC box): 5'-TTATC(C/A)A(C/A)-3'. DnaA binds to ATP and to acidic  
CC phospholipids (by similarity).  
CC -1- SIMILARITY: Belongs to the dnaA family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE014016; AA026756.1; -;  
DR HAMAP; MF 00377; -; 1.  
DR SMART; SMO0382; AAA; 1.  
DR PROSITE; PS01008; DNAA; FALSE NEG.  
KW DNA replication; DNA-binding; ATP (POTENTIAL).  
FT NP\_BIND 162 169 ATP (POTENTIAL).  
SQ SEQUENCE 457 AA; 53365 MW; B2950813FADD4028 CRC64;  
Query Match 9.7%; Score 128.5; DB 1; Length 457;  
Best Local Similarity 22.3%; Pred. No. 0.0028;  
Matches 67; Conservative 46; Mismatches 105; Indels 83; Gaps 12;  
QY 21 KGHVPELYVDN--NRIRYQ-----CP-----CKIYDEERFAEL 56  
DB 34 KNHILLYVAPNERSFWMKIDNYENLKLKNFCNINTPTMLKICKPILQKFFNEL 93  
QY 57 IYSHHQRDTLANKADITYNHRDLVMAADICTATATNGEQVY----- 103  
DB 94 T-----LKNILNSKL--YVNTKLSNITYSSSEINTVTPONFTGQSNQALFKTYKI 146  
QY 104 -----LYLGPFGKSFILGAIANOLSKKRVSTIYL--PERITLKGGFKD 150  
DB 147 AHNPGKNYENPLFLYKSGSGKTHLHVAANTLTKTKNTKIYINSENFQIMTSLKN 206  
QY 151 GSFEKTLHVRANILMLDDIGAEVTPWYRDEV--IGPLHY-----RMVH 195  
DB 207 NTIEFKKYRVSNTLLIDIGFAYKKSQSELFTPTINALLARNQOIIITSDQFOKIH 266

QY 196 ELPTFSSNPDYSELEHMLATRDGEKTKAARI---ERKVSUSPY---PLSGENFRN 249  
 DB 267 GIEFRLKSRF-----ECGLTIRIDPPDLNRTKILKSHIYDINLSYKAAFIPLAKLKS 321  
 QY 250 N 250  
 DB 322 N 322

RESULT 3  
 YQAM\_BACSU  
 ID YQAM\_BACSU STANDARD; PRT; 313 AA.  
 AC P45910;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical protein yqam.  
 GN YQAM.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / JH642;  
 RX MEDLINE=95219086; PubMed=7704261;  
 RA Takemaru K.-I., Mizuno M., Sato T., Takeuchi M., Kobayashi Y.,  
 RT "Complete nucleotide sequence of a skin element excised by DNA  
 rearrangement during sporulation in *Bacillus subtilis*.";  
 RL Microbiology 141:323-327(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / JH642;  
 RX MEDLINE=97124195; PubMed=8969508;  
 RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,  
 RA Kobayashi Y.,  
 RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of  
 the *Bacillus subtilis* genome containing the skin element and many  
 sporulation genes.";  
 RL Microbiology 142:3103-3111(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunze F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertolo M.G., Beesieres P., Bolotin A., Borchert S.,  
 RA Borstis R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Funa S., Gallizzi A., Galleron N.,  
 RA Ghim S.Y., Glaeser P., Goffeau A., Golightly B.J., Grandi G.,  
 RA Glimp S.G., Guy B.J., Haga K., Haeck J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaere-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Paro V., Pohl T.M., Portetle D., Portollik S., Prescott A.M.,  
 RA Pressac B., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha B., Roche B., Rose M., Sadale Y.,  
 RA Sato T., Schanli B., Schleich S., Schroeter R., Scotfene F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takehashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vanier P., Vassarotti A.,  
 RA Viari A., Wambut R., Wedler E., Wedler H., Weltensgiger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,  
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*

RT subtilis.";  
 RL Nature 390:249-256(1997).  
 RN [4]  
 RP IDENTIFICATION.  
 RX MEDLINE=96084975; PubMed=7489895;  
 RA Medigue C., Moszer I., Viari A., Danchin A.,  
 RT "Analysis of a *Bacillus subtilis* genome fragment using a co-operative  
 RT computer system prototype.";  
 RL Gene 165:GC37-GC51(1995)  
 CC -1- SIMILARITY: TO B.SUBTILIS YQXC AND T.HYDROXYENTERIAE HEMOLYSIN  
 CC TLVA.  
 CC CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D32436; BAA06937.1; -  
 DR EMBL: D84432; BAA12388.1; -  
 DR EMBL: Z99117; CAB14567.1; -  
 DR PIR: D69945; D69945.  
 DR Subtilist; BG11264; yqam.  
 DR InterPro: IPR003593; AAA\_Arase.  
 DR SMART: SM00382; AAA; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 313 AA; 36138 MW; 09208795E310D73A CRC64;

Query Match  
 Best local similarity 33.1%; Score 127.5; DB 1; Length 313;  
 Matches 40; Conservative 19; Mismatches 49; Indels 13; Gaps 6;

QY 98 GEQVKGLVYGPFGTSGFLLGAIANOL-KSKRVSTIYLP--EFIRTLKGGFKDGSFE 154  
 DB 161 GGRQNSIALGQPGSGKTHLLTAIMNNLLIKKSVH--CMYFPVBEQMDLKNF--DNLE 216  
 QY 155 KCLHVRKNTIMDDI-----GAEVTPWVDEVIQPLHTRMHELTPFSSNDYSE 209  
 DB 217 AKLDNRKRVVLFIDLPKINGOPRATDW-QVEQIQSVLRYLNHKLILSSSETLIDE 275

QY 210 L 210  
 DB 276 I 276

RESULT 4  
 XKDC\_BACSU  
 ID XKDC\_BACSU STANDARD; PRT; 266 AA.  
 AC P39782;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Phage-like element PSX protein xkdc.  
 GN XKDC.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / SO113;  
 RX MEDLINE=94364963; PubMed=8083174;  
 RA McDowell G.E., Wood H., Devine K.M., McConnell D.J.,  
 RT "Genetic control of bacterial suicide: regulation of the induction of  
 RT PSX in *Bacillus subtilis*.";  
 RL J. Bacteriol. 176:5820-5830(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX Krogsh S., O'Reilly M., Nolan N., Devine K.M.,  
 RL submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]

RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunze F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Besieres P., Bolotin A., Borcherdt S.,  
 RA Borries R., Boutsier L., Brans A., Braun M., Brigelli S.C., Bron S.,  
 RA Brouillet S., Bruchet C.V., Caldwell B., Capuano N.M., Carter N.M.,  
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Duesterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Ertan K.D., Errington J., Fabre C., Ferrari E., Follmer D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galletton N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 RA Giesepe G., Guy B.J., Haga K., Haelech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holappell S., Hosono S., Hulio M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapius A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudaga B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Portwollik S., Prescott A.M.,  
 RA Presecan E., Pujic P., Putnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tanakoshi A., Tanaka T., Terpetra P., Tognoni A.,  
 RA Tobato V., Uchiyama S., Vandenbol M., Vannier P., Vassartot A.,  
 RA Vlati A., Wambuit R., Wedler H., Weitzenecker T.,  
 RA Winters P., Wiput A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
 RT subtilis".  
 RL Nature 350:249-256(1997).  
 CC - FUNCTION: MAY FUNCTION AS A TRANSCRIPTIONAL ANTITERMINATOR.  
 CC - SIMILARITY: STRONG, TO B.SUBTILIS YQAM.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC  
 CC EMBL; Z34287; CA84044.1; -;  
 DR EMBL; Z70177; CA94054.1; -;  
 DR EMBL; Z99110; CAB13110.1; -;  
 DR PIR; I40411; I40411.  
 DR Subtilisin; BG10996; xkdc.  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR SMART; SM00382; AAA; 1.  
 KW ATP-binding; Complete proteome.  
 FT NP BIND 124 131 ATP (POTENTIAL).  
 FT CONFLICT 239 239 K -> R (IN REF. 1).  
 SQ SEQUENCE 266 AA; 30534 MW; 9929C991E9D55A CRC64;

Query Match 9.6%; Score 127; DB 1; Length 266;  
 Best Local Similarity 31.0%; Pred. No. 0.0019;  
 Matches 39; Conservative 25; Mismatches 44; Indels 18; Gaps 7;

QY 92 CTA--INGROV-----GLVPGFSGTSLGALANQKSKVRSSTIYL--EFIR 142  
 DB 101 CTBYADVADYEQIDCKRNSIALGQPSGKTHLTAANEL--MTCYVVPYVFEVGGFT 159  
 QY 143 TLKGGFSGTSLGALANQKSKVRSSTIYL--EFIR 142  
 DB 160 DLKNDP--ALLENKUNMKQADYVIFIDLPKPVNGKRRATDQOLEQYVS--VIANRYLNHK 216  
 QY 198 PTFPSS 203  
 DB 217 PILLS 222

RESULT 5  
 ID DNA\_CAMJ6 STANDARD; PRT; 440 AA.  
 AC G9PJB0;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Chromosomal replication initiator protein dnaA.  
 GN DnaA OR C00001.  
 OS Campylobacter jejuni.  
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;  
 OC Campylobacteraceae; Campylobacter.  
 OX NCBI\_TaxID=197;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCTC 11168;  
 RX MEDLINE=20150912; PubMed=10688204;  
 RA Parkhill J., Wren B.W., Mungall K., Kettle J.M., Churcher C.,  
 RA Basham D., Chillingworth T., Davies R.M., Feltham T., Holtroyd S.,  
 RA Jais K., Karyshev A.V., Moule S., Pallen M.J., Penn C.W.,  
 RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,  
 RA Whitehead S., Barrett B.G.;  
 RT "The genome sequence of the food-borne pathogen Campylobacter jejuni  
 RT reveals hypervariable sequences."  
 RL Nature 403:665-668(2000).  
 CC - FUNCTION: PLAYS AN IMPORTANT ROLE IN THE INITIATION AND REGULATION  
 CC OF CHROMOSOMAL REPLICATION. BINDS TO THE ORIGIN OF REPLICATION;  
 CC IT BINDS SPECIFICALLY DOUBLE-STRANDED DNA AT A 9 BP CONSENSUS  
 CC (DNA BOX): 5'-TATAC(C/A)A(C/A)-3'. DNA BINDS TO ATP AND TO  
 CC ACIDIC PHOSPHOLIPIDS (BY SIMILARITY).  
 CC - SIMILARITY: Belongs to the dnaA family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC  
 CC EMBL; AL139074; CAB72494.1; -;  
 DR HAMAP; MF 00377; -; 1.  
 DR HAMAP; DB1415; DB1415.  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR InterPro: IPR001957; Bac\_DnaA.  
 DR Pfam; PF00308; bac\_dnaa; 1.  
 DR PRINTS; PR00051; DNAA.  
 DR SMART; SM00382; AAA; 1.  
 DR TIGRFAMs; TIGR00362; DnaA; 1.  
 DR PROSITE; PS01008; DNAA; FALSE NEG.  
 KW DNA replication; DNA-binding; ATP-binding; Complete proteome.  
 FT NP BIND 143 150 ATP (POTENTIAL).  
 SQ SEQUENCE 440 AA; 49686 MW; A3FAB90CCED6BD7 CRC64;

Query Match 9.5%; Score 125.5; DB 1; Length 440;  
 Best Local Similarity 26.8%; Pred. No. 0.0048;  
 Matches 41; Conservative 30; Mismatches 51; Indels 31; Gaps 7;

QY 92 CTAITNGEQVGLY---LVPGFSGTSLGALANQKSKVRSSTIYL--EFIR 145  
 DB 123 CKAIANKDKLGLKYNIPYVPGFSGTSLGALANQKSKVRSSTIYL--EFIR 180  
 QY 146 GKGFGSGTSLGALANQKSKVRSSTIYL--EFIR 145  
 DB 181 SNKNGSLDKFHKYKNCVLLIDYVQFLGKTDKIOE-----FF--F 221  
 QY 206 DYSELNHH---LAMPDGEKTKAARILIRVYS 235  
 DB 222 IFNEIKNDGQIITSDNPNNLKG--ITERLKS 253

RESULT 6

DNAC\_BUCAP STANDARD; PRT; 246 AA.

AC 08KAT9;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DE DNA replication protein dnac.

GN DNAC OR BUSG022.

OS Buchnera aphidicola (subsp. Schizaphis graminum).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Buchnera.

NCBI\_TaxID=98794;

SEQUENCE FROM N.A.

MDLINE=2084549; PubMed=12089438;

RA Tamas I., Klaesson L., Candaeck B., Naeslund A.K., Eriksson A.-S., Werngreen J.J., Sandstroem J.P., Moran N.A., Andersson S.G.B.; "50 million years of genomic stasis in endosymbiotic bacteria.";

RT Science 296:2376-2379(2002).

CC -1- FUNCTION: THIS PROTEIN IS REQUIRED FOR CHROMOSOMAL REPLICATION. IT FORMS, IN CONCERT WITH DNAB PROTEIN AND OTHER PREPIMING PROTEINS DNAT, N, N', N'' A PREPIMING PROTEIN COMPLEX ON THE SPECIFIC SITE OF THE TEMPLATE DNA RECOGNIZED BY PROTEIN N' (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE DNAC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AB014077; AAM67594.1; Complete proteome.

KW DNA replication; Primosome; PROBABLY INVOLVED IN THE INTERACTION WITH THE DNAB PROTEIN (BY SIMILARITY).

FT SITE 69

FT SEQUENCE 246 AA; 28497 MW; BF7E2A9BC2ADB5D CRC64;

Query Match 9.2%; Score 122; DB 1; Length 246; Best Local Similarity 23.6%; Pred. No. 0.0046; Matches 48; Conservative 39; Mismatches 86; Indels 30; Gaps 7;

QY 48 DEERFAELITSH---WQRTLNANKDQIYNN-----HRRLDVYMAADICRA 94

DB 35 EGRLSSESLREKNAKMQVLRSGIRLYNMCSPENYKIEHDQRTKLAKKRYAEE 94

QY 95 ITNGEONVGLIYGPRTGKSFILGAIANOLKSKVRSSTIYLPFRITLKGFGKDS-- 152

DB 95 F--NENIASPIFGSGRTGKGNLASAIGNVLIHGKSLIVYADLMSNKKGTFSGTSNI 152

QY 153 -PEKKLHVRVREANITLMDDIGAEVTPWVDEVIQPLHRYMVELPTFPSSNPDYSELE 211

DB 153 TEENLLHNSVLDLMDIDEGMTESRYEK-VIIINOIVDRSSSKSTGMLSLDRGKM 211

QY 212 HHLAMTRDGEKTKARITERRK 234

DB 212 NL-----GS-----VIDRMR 223

RESULT 7

DNAC\_BUCAT STANDARD; PRT; 246 AA.

ID DNAC\_BUCAT

AC P57134;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE DNA replication protein dnac.

GN DNAC OR BU021.

OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum symbiotic bacterium).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Buchnera.

NCBI\_TaxID=118099;

AC SEQUENCE FROM N.A.

RC STRAIN=Tokyo 1998;

RX MEDLINE=20445173; PubMed=10993077;

RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.; "Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS.";

RT Nature 407:81-86(2000).

CC -1- FUNCTION: THIS PROTEIN IS REQUIRED FOR CHROMOSOMAL REPLICATION. IT FORMS, IN CONCERT WITH DNAB PROTEIN AND OTHER PREPIMING PROTEINS DNAT, N, N', N'' A PREPIMING PROTEIN COMPLEX ON THE SPECIFIC SITE OF THE TEMPLATE DNA RECOGNIZED BY PROTEIN N' (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE DNAC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AB001118; BAB12748.1; Complete proteome.

KW DNA replication; Primosome; PROBABLY INVOLVED IN THE INTERACTION WITH THE DNAB PROTEIN (BY SIMILARITY).

FT SITE 69

FT SEQUENCE 246 AA; 28444 MW; C2E56A3BCD1A68F CRC64;

Query Match 8.7%; Score 115.5; DB 1; Length 246; Best Local Similarity 22.5%; Pred. No. 0.016; Matches 49; Conservative 40; Mismatches 88; Indels 41; Gaps 8;

QY 44 KIKYDEE-----RFEALITSH---WQRTLNANKDQIYNN-----HR 79

DB 20 KEKFDNEDILLAMNOEGRLSSESLIREKNAKMQVLRSGIRLYNMCSPDNYKIEHD 79

QY 80 DLDVYMAADICITATNGEONVGLIYGPRTGKSFILGAIANOLKSKVRSSTIYLPB 139

DB 80 GGRKVLKSKRYAEE--NENIASPIFGSGRTGKGNLASAIGNVLIHGKSLIVYAD 137

QY 140 FRTLLKGFGKDS---FEKKLHVRVREANITLMDDIGAEVTPWVDEVIQPLHRYMVE 196

DB 138 LMSNKKGTSGTSNTEENLHLDLSSVDLMDIDEGMTESRYEK-VIIINOIVDRSSSK 196

QY 197 LPTFPSSNPDYSELEHHLAMTRDGEKTKARITERRK 234

DB 197 RSTGMLSNLDHGMKSL-----GS-----VIDRMR 223

RESULT 8

DNA\_AQUAE STANDARD; PRT; 399 AA.

ID DNA\_AQUAE

AC O66659;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Chromosomal replication initiator protein dnaa.

GN DNA OR AQ\_322.

OS Aquifex aeolicus.

OC Bacteria; Aquificales; Aquificaceae; Aquifex.

NCBI\_TaxID=63363;

SEQUENCE FROM N.A.

RP STRAIN=VF5;

RX MEDLINE=9819666; PubMed=9537320;

RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L., Graham D.B., Overbeek R., Sneed M.A., Keller M., Anjay M., Huber R., Feldman R.A., Short J.M., Olson G.J., Swanson R.V.; "The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.";

RT Nature 392:353-358(1998).

RL

RN [2]

RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 76-399.  
 RX MEDLINE=22220049; PubMed=12234917;  
 RA Erzberger J.P., Pirruccello M.M., Berger J.M.;  
 RT "The structure of bacterial DnaA: implications for general mechanisms  
 underlying DNA replication initiation.";  
 RL EMBO J. 21:4763-4773(2002).  
 CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE INITIATION AND REGULATION  
 OF CHROMOSOMAL REPLICATION. BINDS TO THE ORIGIN OF REPLICATION;  
 CC IT BINDS SPECIFICALLY DOUBLE-STRANDED DNA AT A 9 BP CONSENSUS  
 CC (DNAA BOX): 5'-TTATC(C/A)A(C/A)-3'. DNAA BINDS TO ATP AND TO  
 CC ACIDIC PHOSPHOLIPIDS (BY SIMILARITY).  
 CC -1- SIMILARITY: Belongs to the dnaA family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: A800683; AAC06612.1; -  
 CC PIR: G70328; G70328.  
 CC PDB: 1LBQ; 25-SEP-02.  
 CC HAMAP: MF\_00377; -1.  
 CC InterPro: IPR003593; AAA\_ATPase.  
 CC InterPro: IPR001957; Bac DnaA.  
 CC Pfam: PF00308; bac\_dnaA; 1.  
 CC PRINTS: PR00051; DNAA.  
 CC SMART: SM00382; AAA; 1.  
 CC TIGRFAMs: TIGR00362; DnaA; 1.  
 CC PROSITE: PS01008; DNAA; 1.  
 CC DNA replication; DNA-binding; ATP-binding; Complete proteome;  
 CC 3D-structure.  
 FT NP\_BIND 119 126 ATP (POTENTIAL).  
 SQ SEQUENCE 399 AA; 46840 MW; C641CB199AA79906 CRC64;  
 Query Match 8.7%; Score 115; DB 1; Length 399;  
 Best Local Similarity 25.2%; Pred. No. 0.031;  
 Matches 33; Conservative 31; Mismatches 53; Indels 14; Gaps 4;  
 QY 44 KIRYDEERFAELITSHHQRDTLNAK-LKDIYNNRDLVAMAAADDICTAITN-GEQ 100  
 DB 63 EYKEKEKKKVEI-----KDFLNPKYLTLENFVGGNRL---AYEVKALENLGSL 111  
 QY 101 VKGLIYVPGPGTGSFTLGAIANQLKSKVRSTIIVPEFIRTLKGFPGSGSEKQIHRV 160  
 DB 112 YNPIFYGVGTGKTHLQAAGNEAKRGYRVYSSADPQAQVHEHLKKGITNEFRMY 171  
 QY 161 REANIILMLDDI 171  
 DB 172 KSYVDLLLDV 182  
 RESULT 9  
 ISTD\_PSEAE STANDARD; PRT; 265 AA.  
 AC P15026;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Insertion sequence IS21 putative ATP-binding protein.  
 GN ISTD.  
 OS Pseudomonas aeruginosa.  
 OC Plasmid R68.45.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 OC NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99218951; PubMed=2540414;  
 RA Reilmann C., Moore R., Little S., Savioz A., Willetts N.S., Haas D.;  
 CC "Genetic structure, function and regulation of the transposable

RT element IS21.";  
 RL Mol. Gen. Genet. 215:416-424(1989).  
 RN [2]  
 RP REVISION TO 283.  
 RA Berger B.;  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: THE ISTD PROTEIN, ONE OF TWO PROTEINS EXPRESSED ONLY  
 CC WHEN THERE IS A TANDEM REPEAT OF THE IS21 INSERTION SEQUENCE, IS  
 CC NECESSARY FOR THE TRANSPOSITION OF PLASMIDS WITH THAT TANDEM  
 CC REPEAT.  
 CC -1- SIMILARITY: BELONGS TO THE IS21/IS1162 PUTATIVE ATP-BINDING  
 CC PROTEIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X14793; CAA32899.2; -  
 CC InterPro: IPR003593; AAA\_ATPase.  
 CC InterPro: IPR002611; Istd\_ATPbind.  
 CC Pfam: PF01695; Istd; 1.  
 CC SMART: SM00382; AAA; 1.  
 CC Plasmid; Transposable element; ATP-binding.  
 FT NP\_BIND 104 111 ATP (POTENTIAL).  
 SQ SEQUENCE 265 AA; 30528 MW; 904C7ADC5E3A12A9 CRC64;  
 Query Match 8.6%; Score 114; DB 1; Length 265;  
 Best Local Similarity 26.1%; Pred. No. 0.023;  
 Matches 42; Conservative 36; Mismatches 53; Indels 30; Gaps 8;  
 QY 99 EYKGLIYVPGPGTGSFTLGAIANQLKSKVRSTIIVP-EFIRTLKGFPGSGFEKK 156  
 DB 95 ERSSENVILIGPPGVGKTHL-AIALGVKAVDAGHRYLFWPLRLITLTKAKQENLERO 152  
 QY 157 IHRVRANIMLDDIGAEVTPWVRDEVIGPLHYRMV---ELTFPSSN---FDYSE 209  
 DB 153 LQQLSYARVLLIDELG--YLPNNNEA--SLFFGLNRRYEKASTIILTSNGFADWG 206  
 QY 210 L-EHILAMTRDGEKTKAARIIRKYSISTYFVLSGENFR 248  
 DB 207 MRGDHVL-----ATAILDLRHHSITLINIKGESYR 236  
 RESULT 10  
 DNAA\_ACHLA STANDARD; PRT; 445 AA.  
 ID DNAA\_ACHLA  
 AC O9KH08;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Chromosomal replication initiator protein dnaA.  
 GN DNAA.  
 OS Acholeplasma laidlawii.  
 OC Bacteria; Firmicutes; Mollicutes; Acholeplasmatales;  
 OC Acholeplasmataceae; Acholeplasma.  
 OC NCBI\_TaxID=2148;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PG-8B;  
 RX MEDLINE=20242229; PubMed=10779957;  
 RA Taganov K.D., Gushchin A.E., Akopian T.A., Oparina N.Y.,  
 RA Abramychcheva N.Y., Govorun V.M.;  
 RT "Analysis of genes, coding for DNA gyrase from the mycoplasma  
 RT Acholeplasma laidlawii PG-8B.";  
 RL Mol. Biol. (Mosk) 34:292-299(2000).  
 CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE INITIATION AND REGULATION  
 CC OF CHROMOSOMAL REPLICATION. BINDS TO THE ORIGIN OF REPLICATION;  
 CC IT BINDS SPECIFICALLY DOUBLE-STRANDED DNA AT A 9 BP CONSENSUS  
 CC (DNAA BOX): 5'-TTATC(C/A)A(C/A)-3'. DNAA BINDS TO ATP AND TO

CC ACIDIC PHOSPHOLIPIDS (BY SIMILARITY).  
 CC -1- SIMILARITY: Belongs to the dnaa family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF248639; AAF75986.1; -  
 CC DR HAMAP; MF\_00377; -; 1.  
 CC DR InterPro; IPR003593; AAA\_ATPase.  
 CC DR InterPro; IPR001957; Bac\_Dnaa.  
 CC DR Pfam; PF00308; Bac\_dnaa; 1.  
 CC DR PRINTS; PR00051; DNAA.  
 CC DR SMART; SM00382; AAA; 1.  
 CC DR TIGRFAm; TIGR00362; Dnaa; 1.  
 CC DR PROSITE; PS01008; DNAA; 1.  
 CC DR DNA replication; DNA-binding; ATP-binding  
 CC NP\_BIND 144 151 ATP (POTENTIAL).  
 CC FT SEQUENCE 445 AA; 51046 MW; A6045EE53F8C9DB3 CRC64;  
 SQ  
 Query Match 8.4%; Score 111.5; DB 1; Length 445;  
 Best Local Similarity 21.8%; Pred. No. 0.069;  
 Matches 37; Conservative 34; Mismatches 74; Indels 25; Gaps 7;  
 QY 19 EVKGVPELYVDN-NRIKIRYLOCPCKIKY-DEERPEALITSHMORDTLNKLDIY- 75  
 DB 50 FLKRIKLYIAIKINELATKYSSTPVRLKFSQEEVIEEVAADRIKLTIDYRQNLNSTYT 109  
 QY 76 -----MNRDLVDVAMADDICTAITNGEQVKGILYVPEFGTGSFIIIGAIANOLKSK 128  
 DB 110 PSEFVVGKSMFAPRMKVAHPGAVAN-----PFIIFGDELGKTHLMQAGNYILDN 164  
 QY 129 KYASTIIV-----LPEFIRTL-KGSGFGSGFEKTLHRYEANTIMLDDI 171  
 DB 165 DVAKRLIYKADNFIEDFVLSRNKNTKEFNK---YQDIDIVLVDI 211

RESULT 11  
 ID ISTR\_BURCE STANDARD; PRT; 229 AA.  
 AC P55923;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Insertion sequence IS408 putative ATP-binding protein (ORF2).  
 OS Burkholderia cepacia (Pseudomonas cepacia).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Burkholderiaceae; Burkholderia.  
 OX NCBI\_TaxID=292;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 17616 / 249;  
 RX MEDLINE=94302134; PubMed=7518087;  
 RA Byrne A.M., Leesie T.G.;  
 RT "Characteristics of IS401, a new member of the IS3 family implicated  
 RT in plasmid rearrangements in Pseudomonas cepacia.";  
 RL Plasmid 31:138-147(1994).  
 CC -1- SIMILARITY: BELONGS TO THE IS21/IS1162 PUTATIVE ATP-BINDING  
 CC PROTEIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L09108; -; NOT\_ANNOTATED\_CDS.

DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR002611; IserB\_ATPbind.  
 DR Pfam; PF01695; IserB; 1.  
 DR SMART; SM00382; AAA; 1.  
 KW Transposable element; ATP-binding.  
 FT NP\_BIND 105 112 ATP (POTENTIAL).  
 SQ SEQUENCE 229 AA; 25929 MW; ACSD619C729163E1 CRC64;  
 Query Match 8.2%; Score 109.5; DB 1; Length 229;  
 Best Local Similarity 27.0%; Pred. No. 0.045;  
 Matches 50; Conservative 32; Mismatches 80; Indels 23; Gaps 8;  
 QY 59 SHMORDTLNATLK-----DIYNNHRDLR-----VAMADDICTAITNGEQVKGILYXG 108  
 DB 51 TRRLRLRAALTKRQPCVEDIHYRQTRGLDQRYATLAG--CDWVRAQ--NLITLG 105  
 QY 109 PFGTGSFIIIGAIANOLSKKVRSTIIVPEFIRTLKGSGFEKTLHRYEANTIML 168  
 DB 106 PFGAGKTWLACAFGQACQSGSVFVYVARLFEELKIHGSGSFTRRLAQIAKIDVLIL 165  
 QY 169 DDTGAEVTPWVRDVIQGLHYRWVHELPTFFSSNPDQSELEHIAWTRDDEEYTKAR 228  
 DB 166 DDMGLQDLQQAARNDL-EVLDDR-VGTRSTVITSQL--PLEHWAWLQD--PTLADA 217  
 QY 229 IIERV 233  
 DB 218 IIDRL 222

RESULT 12  
 ID DNAA\_PASMU STANDARD; PRT; 451 AA.  
 AC Q9CQJ4;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Chromosomal replication initiator protein dnaa.  
 GN DNAA OR PM1161.  
 OS Pasteurella multocida.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Pasteurella.  
 OX NCBI\_TaxID=747;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Pm70;  
 RX MEDLINE=21145866; PubMed=11248100;  
 RA May B.U., Zhang Q., Li L.L., Paustian M.L., Whitlam T.S., Kapur V.;  
 RT "Complete genomic sequence of Pasteurella multocida Pm70.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).  
 CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE INITIATION AND REGULATION  
 CC OF CHROMOSOMAL REPLICATION. BINDS TO THE ORIGIN OF REPLICATION  
 CC IT BINDS SPECIFICALLY DOUBLE-STRANDED DNA AT A 9 BP CONSENSUS  
 CC (DNAA BOX): 5'-TTATC(C/A)A(C/A)-3'. DNAA BINDS TO ATP AND TO  
 CC ACIDIC PHOSPHOLIPIDS (BY SIMILARITY).  
 CC -----  
 CC -1- SIMILARITY: Belongs to the dnaa family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AE006156; AAK03245.1; -  
 CC DR HAMAP; MF\_00377; -; 1.  
 CC DR InterPro; IPR003593; AAA\_ATPase.  
 CC DR InterPro; IPR001957; Bac\_Dnaa.  
 CC DR Pfam; PF00308; Bac\_dnaa; 1.  
 CC DR PRINTS; PR00051; DNAA.  
 CC DR SMART; SM00382; AAA; 1.  
 CC DR TIGRFAm; TIGR00362; Dnaa; 1.  
 CC DR PROSITE; PS01008; DNAA; 1.

KM DNA replication; DNA-binding; ATP-binding; Complete proteome.  
 FT NP BIND 156 ATP (POTENTIAL).  
 SQ SEQUENCE 451 AA; 51378 MW; 76A0ACD237FF2 CRC64;

Query Match 8.1%; Score 107; DB 1; Length 451;  
 Best Local Similarity 21.3%; Pred. No. 0.16;  
 Matches 52; Conservative 47; Mismatches 93; Indels 52; Gaps 12;

QY 19 FVKGHVELYV-----DNNRIKIRYLQCPCKIKYDEERFEALITSHHMORDTL 67  
 DB 50 FIKGVDESYLAQITKVAQELSGNAELV---QKGVKPEPKQAQSAFTHNKSEBK 106  
 QY 68 NAKLKDIYMNDRDL--VMAADICTAINT-----GEOVK--LYLGGPFGKSEFYL 118  
 DB 107 PQVIRSYLNPKHVFNPEVKGSKNQALARAQVADNPGEPSNPLFLYGSTGLGKTHLL 166  
 QY 119 GATNOLKSKKVRSTIYL--PEFIRTLKGGFKGSGRE--KKLRVREANIMLDDT--- 171  
 DB 167 HAINGNLSRNTNARVLYIHANNFMQMVANVRDNKDEKFKYRSIDA--LVDDIQPF 224  
 QY 172 GAEVTPWVNDVIGPLTHYRVHDELPTFSSNPDYSLEHHLAMTRDGEKTKARIIIE 231  
 DB 225 AEKKTQ-----EEFRIHFNMLPDTG---RQITLTSRTPK-EIEKUEB 264  
 QY 232 RVKS 235  
 DB 265 RLKS 268

RESULT 13  
 ID ISTB\_BACST STANDARD; PRT; 251 AA.  
 AC Q45619;  
 DT 01-NOV-1997 (Rel. 35, Last Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Insertion sequence IS5376 putative ATP-binding protein.  
 OS Bacillus stearothermophilus.  
 CC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.  
 NC NCBI\_TaxID=1422;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CU21;  
 RX MEDLINE=93181468; PubMed=8382825;  
 RA Xu K., He Z.-Q., Mao Y.-M., Shen R.-Q., Sheng Z.-J.;  
 RT "On two transposable elements from Bacillus stearothermophilus.";  
 RL Plasmid 29.1-9(1993).  
 CC -1- SIMILARITY: BELONGS TO THE IS21/IS1162 PUTATIVE ATP-BINDING PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X67861, CAA48046.1, -.  
 DR PIR: S23889;  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR InterPro: IPR002611; IctB\_ATPbind.  
 DR Pfam: PF01695; IctB; 1.  
 DR SMART: SM00382; AAA; 1.  
 KM Transposable element; ATP-binding.  
 FT NP\_BIND 105 112  
 SQ SEQUENCE 251 AA; 29286 MW; FA8FP4F55B2FB527 CRC64;

Query Match 8.0%; Score 106.5; DB 1; Length 251;  
 Best Local Similarity 22.1%; Pred. No. 0.089;  
 Matches 49; Conservative 40; Mismatches 98; Indels 35; Gaps 8;

QY 45 IKYDERF---EALITSHHMORDTLNAKLDIYMNDRDLVAM--AAD--ICTAINT 97

DB 33 ISVSEFLRLEAEIVKQARSIGTLIKLSKLPYRKTIIDFDFAGSVDERRIEILLTL 92  
 QY 98 G--EQVKGILYVPGFGKSPFIIAGTANOLKSKKVRSTIYLBFRITLKGFPDGSFEK 155  
 DB 93 SFIDRKNILFLDPGGIKTHLALISGWEIAGYKTFITADLVNQLRAQEGLEK 152  
 QY 156 KLRVREANIMLDDIGAEVTPWVNDVIGPLHY-----RMVHELPTFSSNPDYSR 209  
 DB 153 KLRVFPVPTVLIIDEMKYLKDP-----NSAHYFQVIARVYHAPLIILSNKSGE 204  
 QY 210 LEHHLAMTRDGE--EKTAKARIIRVKSLSSTPYFSGENR 248  
 DB 205 W-----GEIVGDSVLATAMDLRLHSHIIFNLKGESEYR 237

RESULT 14  
 ID ISTB\_PSEFL STANDARD; PRT; 231 AA.  
 AC Q51762;  
 DT 01-NOV-1997 (Rel. 35, Last Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Insertion sequence IS1162 putative ATP-binding protein.  
 OS Pseudomonas fluorescens.  
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.  
 NC NCBI\_TaxID=294;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ST;  
 RX MEDLINE=95212933; PubMed=7698671;  
 RA Solinas F., Marconi A.M., Ruzzi M., Zennaro B.;  
 RT "Characterization and sequence of a novel insertion sequence, IS1162, from Pseudomonas fluorescens.";  
 RL Gene 155:77-82(1995).  
 CC -1- SIMILARITY: BELONGS TO THE IS21/IS1162 PUTATIVE ATP-BINDING PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X79443; CAA55960.1, -.  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR InterPro: IPR002611; IctB\_ATPbind.  
 DR Pfam: PF01695; IctB; 1.  
 DR SMART: SM00382; AAA; 1.  
 KM Transposable element; ATP-binding.  
 FT NP\_BIND 107 114  
 SQ SEQUENCE 231 AA; 25997 MW; 341592AEA7D794A CRC64;

Query Match 7.9%; Score 105.5; DB 1; Length 231;  
 Best Local Similarity 21.4%; Pred. No. 0.097;  
 Matches 33; Conservative 33; Mismatches 61; Indels 27; Gaps 3;

QY 35 KIRYLQCPCKIKYDEERFEALITSHHMORDTLNAKLDIYMNDRDLVAMAAADICTA 94  
 DB 55 RLRLKQKQKTKY-----ASACLEDIDRRRGRGLTDLRLTA-----T 90  
 QY 95 ITNGEYK---GLYVPGFGKSPFIIAGTANOLKSKKVRSTIYLBFRITLKGFGKOG 151  
 DB 91 LAGSDIRQRHNLILGPGVGKTFWPCALGHCACQGSATVLRTPRLLEQRIAHGDG 150  
 QY 152 SPEKKLRVREANIMLDDIGAEVTPWVNDVY 185  
 DB 151 SFGRITQQLAKVDVILTDWGLAALAEENARHLL 184





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 15:11:40 ; Search time 28.8858 Seconds  
(without alignments)  
2233.384 Million cell updates/sec

Title: US-09-689-952-18

Perfect score: 1328

Sequence: 1 YKDQKXHDGKHFADCPNFV.....ERVKSLSPTPYFLSGNFNN 250

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp Vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1328	100.0	306	16	Q99TH8
2	1320	99.4	306	16	Q8NW67
3	1189	89.5	306	16	Q8CNV2
4	531.5	40.0	311	16	Q9K863
5	523	39.4	308	16	Q8CX80
6	475.5	35.8	307	16	Q92BF4
7	472.5	35.6	307	16	Q8Y6X1
8	397.5	29.9	293	16	Q8B2J5
9	397.5	29.9	300	16	Q9A1D1
10	392.5	29.6	299	16	Q8CWM7
11	381	28.7	298	16	Q8CWP7
12	365	27.5	300	16	Q8S3T8
13	361	27.2	298	16	Q8DY72
14	342	25.8	293	16	Q97PC7
15	342	25.8	293	16	Q9CHH9
16	205.5	15.5	74	2	Q05653

17	194	14.6	343	16	Q8EWK0	Q8EWK0 mycoplasma
18	173	13.0	297	2	Q48991	Q48991 mycoplasma
19	157	11.8	209	2	Q48F03	Q48F03 mycoplasma
20	153	11.5	259	9	Q96022	Q96022 bacteriophage
21	151	11.4	259	9	Q9B0F8	Q9B0F8 staphylococ
22	145	10.9	261	16	Q932A3	Q932A3 staphylococ
23	145	10.9	282	16	Q97H85	Q97H85 clostridium
24	140	10.5	257	9	Q9BHR8	Q9BHR8 staphylococ
25	140	10.5	316	16	Q9P014	Q9P014 ureaplasma
26	138.5	10.4	235	16	Q67056	Q67056 aquifex aeo
27	129	9.7	285	16	Q8E256	Q8E256 leptospira
28	127	9.6	249	2	Q93SD5	Q93SD5 escherichia
29	120.5	9.1	316	2	P95446	P95446 pseudomonas
30	114	8.6	457	16	Q8XPG2	Q8XPG2 clostridium
31	113.5	8.5	195	17	Q8PTM8	Q8PTM8 methanobact
32	113.5	8.5	257	17	Q8TWM8	Q8TWM8 methanobact
33	113.5	8.5	257	17	Q8TH29	Q8TH29 methanobact
34	113	8.5	327	16	Q8R6U1	Q8R6U1 thermomanae
35	113	8.5	796	17	Q8TZV0	Q8TZV0 pyrococcus
36	112.5	8.5	261	2	Q8G101	Q8G101 pseudomonas
37	112	8.4	287	16	Q926A4	Q926A4 listeria in
38	112	8.4	291	16	Q9C1B5	Q9C1B5 lactococcus
39	112	8.4	294	9	Q8L7N1	Q8L7N1 lactococcus
40	112	8.4	294	9	Q9AZP7	Q9AZP7 bacteriophage
41	112	8.4	520	17	Q26312	Q26312 methanobact
42	111.5	8.4	296	9	Q38098	Q38098 bacteriophage
43	111.5	8.4	475	10	Q9LPI1	Q9LPI1 arabidopsis
44	111	8.4	262	2	Q93P71	Q93P71 microscilla
45	111	8.4	350	16	Q921Y8	Q921Y8 rickettsia

#### ALIGNMENTS

##### RESULT 1

ID Q99TH8 PRELIMINARY: PRT: 306 AA.  
AC Q99TH8:  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE Primosomal protein.  
GN DNAI OR SAV1684 OR SA1507.  
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and  
OS Staphylococcus aureus (strain N315).  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_Taxid=158878, 158879;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECTRES-S.aureus (strain Mu50), and S.aureus (strain N315).  
RX MEDLINE=2111952; PubMed=11418146;  
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Itoh T., Ito T.,  
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hoshoyama A.,  
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,  
RA Mizutani-Uji Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
RA Kanihisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;  
RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
RT aureus.";  
RT Lancet 357:1225-1240(2001).  
DR EMBL; AF003363; BAB57846.1; -;  
DR EMBL; AF003134; BAB42774.1; -;  
KW Complete proteome.  
SQ SEQUENCE 306 AA; 35635 MW; 7F3440E89643505B CRC64;

Query Match 100.0%; Score 1328; DB 16; Length 306;  
Best Local Similarity 100.0%; Pred. No. 7.3e-107;  
Matches 250; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 YKDQKXHDGKHFADCPNFVKGVPELVYDNNRIKIRYLQCCCKIKYDEREFAELITSH 60  
DB 57 YKDQKXHDGKHFADCPNFVKGVPELVYDNNRIKIRYLQCCCKIKYDEREFAELITSH 116

QY 61 HMORDTLNAKLKDIYNNHRDLVAMAAADICTAINTGEOVKGLYLGPFGTGSFLGA 120  
 DB 117 HMORDTLNAKLKDIYNNHRDLVAMAAADICTAINTGEOVKGLYLGPFGTGSFLGA 176  
 QY 121 INANQSKKVRSTIYLPFRITLKGFGKGSFEKLLHVRREANIIMLDDIGAEVTPWV 180  
 DB 177 INANQSKKVRSTIYLPFRITLKGFGKGSFEKLLHVRREANIIMLDDIGAEVTPWV 236  
 QY 181 RDEVIGPLHYRWVHLPFPSSNFDYSELEHLLAMTRDGEKTKAARIIRKXSLSTPY 240  
 DB 237 RDEVIGPLHYRWVHLPFPSSNFDYSELEHLLAMTRDGEKTKAARIIRKXSLSTPY 296  
 QY 241 FLGSENFRRN 250  
 DB 297 FLGSENFRRN 306

## RESULT 2

Q8NM67 PRELIMINARY; PRT; 306 AA.  
 AC Q8NM67;  
 DT 01-OCT-2002 (TREMBlrel. 22, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE Primosomal protein.  
 GN DNAI OR MM1627.  
 OS Staphylococcus aureus (strain MW2).  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OX NCBI TaxID=196620;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22040717; PubMed=12044378;  
 RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,  
 RA Nagai Y., Iwano N., Asano K., Naito T., Kuroda H., Cui L.,  
 RA Yamamoto K., Hiramatsu K.;  
 RT "Genome and virulence determinants of high virulence community-  
 RT acquired MRSA".  
 RL Lancet 359:1819-1827(2002).  
 DR EMBL; AP004827; BAB95492.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 306 AA; 35626 MW; C0FCA3752D934B4 CRC64;

Query Match 99.4%; Score 1320; DB 16; Length 306;  
 Best Local Similarity 99.6%; Pred. No. 3.6e-106;  
 Matches 249; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YKDQKHVGHKADCPNFKGVHPELYVDNNRIKIRYLOCPCKIKYDERFEALITSH 60  
 DB 57 YKDQKHVGHKADCPNFKGVHPELYVDNNRIKIRYLOCPCKIKYDERFEALITSH 116  
 QY 61 HMORDTLNAKLKDIYNNHRDLVAMAAADICTAINTGEOVKGLYLGPFGTGSFLGA 120  
 DB 117 HMORDTLNAKLKDIYNNHRDLVAMAAADICTAINTGEOVKGLYLGPFGTGSFLGA 176  
 QY 121 INANQSKKVRSTIYLPFRITLKGFGKGSFEKLLHVRREANIIMLDDIGAEVTPWV 180  
 DB 177 INANQSKKVRSTIYLPFRITLKGFGKGSFEKLLHVRREANIIMLDDIGAEVTPWV 236  
 QY 181 RDEVIGPLHYRWVHLPFPSSNFDYSELEHLLAMTRDGEKTKAARIIRKXSLSTPY 240  
 DB 237 RDEVIGPLHYRWVHLPFPSSNFDYSELEHLLAMTRDGEKTKAARIIRKXSLSTPY 296  
 QY 241 FLGSENFRRN 250  
 DB 297 FLGSENFRRN 306

RESULT 3  
 Q8CNY2 PRELIMINARY; PRT; 306 AA.  
 AC Q8CNY2;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Primosomal protein.  
 GN SE1358.  
 OS Staphylococcus epidermidis.  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OX NCBI TaxID=1282;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 12228;  
 RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,  
 RA Chen Z., Wen Y.;  
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE016748; AAC04957.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 306 AA; 35736 MW; EDD9F7B52DBFCA16 CRC64;

Query Match 89.5%; Score 1189; DB 16; Length 306;  
 Best Local Similarity 87.2%; Pred. No. 7.9e-95;  
 Matches 218; Conservative 23; Mismatches 9; Indels 0; Gaps 0;

QY 1 YKDQKHVGHKADCPNFKGVHPELYVDNNRIKIRYLOCPCKIKYDERFEALITSH 60  
 DB 57 YKDQKHVGHKADCPNFKGVHPELYVDNNRIKIRYLOCPCKIKYDERFEALITSH 116  
 QY 61 HMORDTLNAKLKDIYNNHRDLVAMAAADICTAINTGEOVKGLYLGPFGTGSFLGA 120  
 DB 117 HMORDTLNAKLKDIYNNHRDLVAMAAADICTAINTGEOVKGLYLGPFGTGSFLGA 176  
 QY 121 INANQSKKVRSTIYLPFRITLKGFGKGSFEKLLHVRREANIIMLDDIGAEVTPWV 180  
 DB 177 INANQSKKVRSTIYLPFRITLKGFGKGSFEKLLHVRREANIIMLDDIGAEVTPWV 236  
 QY 181 RDEVIGPLHYRWVHLPFPSSNFDYSELEHLLAMTRDGEKTKAARIIRKXSLSTPY 240  
 DB 237 RDEVIGPLHYRWVHLPFPSSNFDYSELEHLLAMTRDGEKTKAARIIRKXSLSTPY 296  
 QY 241 FLGSENFRRN 250  
 DB 297 YLTGKNFRRN 306

## RESULT 4

Q9K863 PRELIMINARY; PRT; 311 AA.  
 AC Q9K863;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Primosome component (Helicase loader).  
 GN DNAI OR BH3144.  
 OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI TaxID=8665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RX MEDLINE=20512582; PubMed=11058132;  
 RA Takai H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
 RA Fujii F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
 RA Horikoshi K.;  
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
 RT halodurans and genomic sequence comparison with Bacillus subtilis".  
 RT Nucleic Acids Res. 28:4317-4331(2000).  
 RL EMBL; AP001517; BAB06863.1; -.  
 DR InterPro; IPR003593; AAA\_Atpase.  
 DR SMART; SM00382; AAA; 1.  
 DR Helicase; Complete proteome.  
 SQ SEQUENCE 311 AA; 35822 MW; 6AFB3073431363BB CRC64;

Query Match 40.0%; Score 531.5; DB 16; Length 311;  
 Best Local Similarity 41.2%; Pred. No. 6.9e-38;  
 Matches 107; Conservative 55; Mismatches 79; Indels 19; Gaps 6;

```

QY 1 YKDOQKHVDGH--KPADCPNFVKGHVELYVNNRIRKIRYLOCPCKIKYDEERFEALIT 58
DB 59 YQAEQSHC-AHCGLOKCPYLMKGYOPTLYVEDSLSTSPDLKEBEREKKSLSLR 117
QY 59 SHHQBORTLNAKLKDI-----YNNHRDLVAMADDICTAITNGEQVGLYVGPFG 111
DB 118 SLIYPIREILAEKPDVSEBSRGSIAHR-ALBFLASAKP-----GSDGMGLYVKGKG 169
QY 112 TKSFFLIGAIANOLKSKVRSTIYLPFRTILKGFQKSGFEKKLHVRREANIMLDDI 171
DB 170 VGTFTLMLGAIAMELKRGIDSTIVYVDFPRELKOSIGDTPFOCKIDPVKNAQVLLFDI 229
QY 172 GAEKVTPWREDEYIGPLHYRMVHLEPTFSSNPDYSELEHNLAMT-ROGEERTKARII 230
DB 230 GAETMTSWADDVGLYLOTRIMEKLPFTLSNYDIDELSHLAINDKSTELLAKKRV 289
QY 231 ERYKSLSTPYLGSNGFRNN 250
DB 290 ERIRHYTVSMVGGQRREH 309

RESULT 5
Q8CB0 PRELIMINARY; PRT; 308 AA.
AC 08CB0;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Primosome component (Hellicase loader).
GN Dnal OR 082156.
OS Oceanobacillus ihyensia.
OC Bacteria; Firmicutes; Bacillales; Oceanobacillus.
ON NCBI_Taxid=182710;
RX STRAIN=HTEB31 / DSM 14371 / JCM 11309;
RX MEDLINE=2220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RA Genome sequence of Oceanobacillus ihyensia isolated from the Itheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments."
RL Nucleic Acids Res. 30:3927-3935(2002).
DR EMBL: AP004600; BAC14112.1; -.
KW Hellicase; Complete proteome.
SQ SEQUENCE 308 AA; 35729 MW; 8D4C2F70E480AE33 CRC64;

Query Match 39.4%; Score 523; DB 16; Length 308;
Best Local Similarity 42.4%; Pred. No. 3, 7e-37;
Matches 106; Conservative 51; Mismatches 91; Indels 2; Gaps 2;

QY 1 YKDOQKHVDGH--KPADCPNFVKGHVELYVNNRIRKIRYLOCPCKIKYDEERFEALIT 59
DB 60 YKQSOQCRCKSLGSCQNMIOGYSPVLEADNNDIRLSYKCKHREBQNEQOKLID 119
QY 60 HHMQORTLNAKLKDIYNNHRDLVAMADDICTAITNGEQVGLYVGPFGSGKSPITG 119
DB 120 LYMPEKILQARISDVIODEH-RSNALGKVLDFLEASRQELPKGLVLYSGFVGKTYLLG 178
QY 120 AINQKSKKVRSTIYLPFRTILKGFQKSGFEKKLHVRREANIMLDDIGAEEVTP 179
DB 179 AINELKLEYSISLYMPEFVAEIKSSPFQDSFNRKVPFKADILMDMGAEMQSAN 238
QY 180 VRDEVIGPLHYRMVHLEPTFSSNPDYSELEHNLAMTRDEEKTAAARIIERVKSLSTP 239
DB 239 FRDEVIGSVLYQYRMMEGLPVFITSNYDLDQLQELSTTRNGVQVAGRIIERIKQVTTD 298
QY 240 YELSGENFRN 249
DB 299 VKLSGPNRRS 308

RESULT 6

```

```

Q92BF4 PRELIMINARY; PRT; 307 AA.
ID 092BF4;
AC 092BF4;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Primosome component (Hellicase loader) Dnal.
GN Dnal OR L1N1595.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
ON NCBI_Taxid=1642;
RX STRAIN=CLIP 11262 / Serovar 6a;
RX PubMed=11679669;
RA Glaeser P., Frangeul L., Buchrieser C., Rueniock C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chattradorty T.,
RA Charbit A., Chetouani F., Couve E., de Darvar A., Deloux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Duessegret O.,
RA Ertlan K.-D., Fehli H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kuntz F., Kurapkhat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjati H.,
RA Nordieck G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL: AL596169; CAC96826.1; -.
DR Listeria; L1N1595; -.
DR InterPro: IPR003593; AAA_Arpase.
DR SMART: SM00382; AAA; 1.
KW Hellicase; Complete proteome.
SQ SEQUENCE 307 AA; 35988 MW; A8517663BA8F123F CRC64;

Query Match 35.8%; Score 475.5; DB 16; Length 307;
Best Local Similarity 42.6%; Pred. No. 4, 8e-33;
Matches 98; Conservative 45; Mismatches 84; Indels 3; Gaps 2;

QY 22 GHVPELYVNNRIRKIRYLOCPCKIKYDEERFEALITSHHQBORTLNAKLKDIYNNHRDR 81
DB 77 GYAPKVLVNGERITVYVPTKKEIEBKRAVARRIRSLYMKQVVDAMVADFYTDESR 136
QY 82 LVVMAADDICTAI--TNGEQVGLYVGPFGSGKSPITGAIANOLKSKVRSTIYLP 139
DB 137 KIALVEAYOFLNNYPPONERVKGLPFHSGFSGSYLALAKELALGISTITLYLPE 196
QY 140 FIRTILKGFQKSGFEKKLHVRREANIMLDDIGAEEVTPWREDEYIGPLHYRMVHELPT 199
DB 197 FPREYKQSDSDTVGKIKOPAFETEVLMDDIGAESMTMTWTDVYGLALQFRMOBELPT 256

QY 200 FSSNPDYSELEHNLAMTRDG-EKTKAARIIERYKSLSTPYLGSNGFRN 248
DB 257 FSSNPNMQLNHNLMFAQNGTEKIKARIMERYALISKEYVLEBKNRR 306

RESULT 7
Q8Y6X1 PRELIMINARY; PRT; 307 AA.
ID 08Y6X1;
AC 08Y6X1;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-UTN-2002 (TREMBLrel. 21, Last annotation update)
DE Primosome component (Hellicase loader) Dnal.
GN Dnal OR LMO1560.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
ON NCBI_Taxid=1639;
RX STRAIN=BGD-e / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaeser P., Frangeul L., Buchrieser C., Rueniock C., Amend A.,

```

RA Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,  
 RA Charbit A., Checouani F., Couve E., de Darvar A., Dohoux P.,  
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durand L., Duseigneur O.,  
 RA Ercian K.-D., Fehli H., Garcia-del Portillo F., Garrido P.,  
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Haut J., Jackson D.,  
 RA Jones L.W., Kaerst U., Kreft J., Kuhn M., Kunst F., Kuxapkat G.,  
 RA Madueno E., Matoumnan A., Mata Vicente J., Ng E., Nedjati H.,  
 RA Nordstedt G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
 RA Rammel B., Rose M., Schlueder T., Simeos N., Tierrez A.,  
 RA Vazquez-Boland J.-A., Voss H., Wehlend J., Coesart P.,  
 RT "Comparative genomics of *Listeria species*,"  
 RL Science 294:849-852(2001).  
 DR EMBL; AL591979; CAC99638.1; -  
 DR Lactidist; LMO01560; -  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR SMART; SM00382; AAA; 1.  
 KW Helicase; Complete proteome.  
 SQ SEQUENCE 307 AA; 35970 MW; CF2FD9350CEB0F48 CRC64;

Query Match 35.6%; Score 472.5; DB 16; Length 307;  
 Best Local Similarity 42.1%; Pred. No. 8,7e-33;  
 Matches 98; Conservative 47; Mismatches 79; Indels 9; Gaps 3;

QY 22 GHVPELVYDNNRIRKIRYLQPCCKIKYDEERFEALITSHMORDTLNAKLDIYMHRRDR 81  
 DB 77 GYAPKVLNAGEFIVTYPTKEKIEBDRABERRIRSLYMPQVADNLADFTDESR 136  
 QY 82 LDVAMADDICTAITN-----GEQVKGILYGPFGTCKSLTIGAIANOLSKYKSTIY 136  
 DB 137 ---QLALVEYQFLNANNPPKSGRVERKGLFTHGSGTGKSYLLAALAEALAKIGISTLLVY 193  
 QY 137 LPEFIRTLKGGFKDGSPEKTLHVRANILMLDIGAEEVTPWVRDEVIGPLHYRWVHE 196  
 DB 194 LPEFMKRVKSGISDNTYGEIKQPAKREVLMLDDIGESMTANTRDVLAIIQFRNGE 253  
 QY 197 LPTFFSSNFDYSELEHHLAMTRDG-EERKTAARIIEVKSLSTPYLSENGFR 248  
 DB 254 LPTFFSSNFMDDQENHLMFAONGTEKLAARRIMERVRLSKVNLGNRR 306

## RESULT 8

Q8P2J5 PRELIMINARY; PRT; 293 AA.  
 ID Q8P2J5  
 AC Q8P2J5;  
 DT 01-OCT-2002 (TREMBlrel. 22, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 GN Putative primosome component (helicase loader).  
 GN SPYM18\_0332.  
 OS Streptococcus pyogenes (serotype M18).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 NCBI\_TaxID=186103;  
 RN NCBI  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MGAS8232 / serotype M18;  
 RX MEDLINE=21927593; PubMed=11917108;  
 RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,  
 RA Sylva G.L., Studevant D.B., Ricklets S.M., Porcella S.F.,  
 RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,  
 RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;  
 RT "Genome sequence and comparative microarray analysis of serotype M18  
 RT group A Streptococcus strains associated with acute rheumatic fever  
 RT outbreaks,"  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).  
 DR EMBL; AB009978; AAL97087.1; -  
 KW Helicase; Complete proteome; Hypothetical protein.  
 SQ SEQUENCE 293 AA; 33357 MW; 0E87AD94BDB59986 CRC64;

Query Match 29.9%; Score 397.5; DB 16; Length 293;  
 Best Local Similarity 35.9%; Pred. No. 2.5e-26;  
 Matches 84; Conservative 56; Mismatches 91; Indels 3; Gaps 3;

QY 17 PNFV-KGHVPELVYDNNRIRKIRYLQPCCKIKYDEERFEALITSHMORDTLNAKLDIY 75  
 DB 61 PSYIAGYQPIILAMNGYADVSLTEKELVEAQKAAISERIQVLSPSYRHHISDDID 120  
 QY 76 MNRDLDVAMADDICTAITNGEQVKGILYGPFGTCKSLTIGAIANOLSKK-VRSSTI 134  
 DB 121 VNNASMEAFSAIILDFVEQYPSAEQ-KGLYLGDMGIGSYLLAAMAHLSKKGYSTTL 179  
 QY 135 IYLPFIRTLKGGFKDGSPEKTLHVRANILMLDIGAEEVTPWVRDEVIGPLHYRWV 194  
 DB 180 LHPSPALIDVKAISNGSVKSEIDAVKNVPLILDDIGABQATSWRDEVLOVILQYRML 239  
 QY 195 HELPTFFSSNFDYSELEHHLAMTRDGEKTKAARIIEVKSLSTPYLSENGFR 248  
 DB 240 ELPTFTFNISYRADLERKMATIKGSDEYVQAKRWVERRYLAREHLEGANRR 293

## RESULT 9

Q9AID1 PRELIMINARY; PRT; 300 AA.  
 ID Q9AID1  
 AC Q9AID1;  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DE Putative primosome component (helicase loader).  
 GN DNAI OR SPY0340 OR SPYM3\_0248.  
 OS Streptococcus pyogenes, and  
 OS Streptococcus pyogenes (serotype M3).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 NCBI\_TaxID=1314, 198466;  
 RN NCBI  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S. pyogenes; STRAIN=SF370 / ATCC 700294 / Serotype M1;  
 RX MEDLINE=21192684; PubMed=11296296;  
 RA Ferretic J.J., Mcshan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,  
 RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,  
 RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,  
 RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;  
 RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes,"  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S. pyogenes; STRAIN=MGAS315 / Serotype M3;  
 RX MEDLINE=22133808; PubMed=12122206;  
 RA Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,  
 RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,  
 RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,  
 RA Schlievert P.M., Musser J.M.;  
 RT "Genome sequence of a serotype M3 strain of group A Streptococcus:  
 RT phage-encoded toxins, the high-virulence phenotype, and clone  
 RT emergence,"  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).  
 DR EMBL; AB006498; AAK33392.1; -  
 DR EMBL; AE014141; AAM78855.1; -  
 KW Hypothetical protein; Helicase; Complete proteome.  
 SQ SEQUENCE 300 AA; 34146 MW; 489B27A789338C5A CRC64;

Query Match 29.9%; Score 397.5; DB 16; Length 300;  
 Best Local Similarity 35.9%; Pred. No. 2.6e-26;  
 Matches 84; Conservative 56; Mismatches 91; Indels 3; Gaps 3;

QY 17 PNFV-KGHVPELVYDNNRIRKIRYLQPCCKIKYDEERFEALITSHMORDTLNAKLDIY 75  
 DB 68 PSYIAGYQPIILAMNGYADVSLTEKELVEAQKAAISERIQVLSPSYRHHISDDID 127  
 QY 76 MNRDLDVAMADDICTAITNGEQVKGILYGPFGTCKSLTIGAIANOLSKK-VRSSTI 134  
 DB 128 VNNASMEAFSAIILDFVEQYPSAEQ-KGLYLGDMGIGSYLLAAMAHLSKKGYSTTL 186  
 QY 135 IYLPFIRTLKGGFKDGSPEKTLHVRANILMLDIGAEEVTPWVRDEVIGPLHYRWV 194  
 DB 187 LHPSPALIDVKAISNGSVKSEIDAVKNVPLILDDIGABQATSWRDEVLOVILQYRML 246

QY 195 HELPTFFSSNFDYSELEHHLAMTRDGEKTKARITERYKSLTPYFSGENR 248  
 DB 247 BELPTFTSYNSADLERKATTKGSDETWQAKRVMERVAYLAKREPHLBSANRR 300

## RESULT 10

Q8CWP7 PRELIMINARY; PRT; 299 AA.

AC 08CWP7; 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DE Putative DNA replication protein, primosome component (Hellicase loader)  
 GN DNAI OR SMU.1921.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 NC Streptococcus.  
 NCBI\_Taxid=1309;

RP SEQUENCE FROM N.A.  
 RC STRAIN=UAI59 / ATCC 700610 / Serotype C;  
 RX MEDLINE=2295063; PubMed=12397186;  
 RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,  
 RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,  
 RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.,  
 RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental  
 pathogen."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439 (2002).  
 DR EMBL; AB015016; AAN59532.1; -  
 KW Hellicase: Complete proteome.  
 SQ SEQUENCE 299 AA; 34410 MW; 98CAAB56F5524BD1 CRC64;

Query Match 29.6%; Score 392.5; DB 16; Length 299;  
 Best Local Similarity 34.5%; Pred. No. 7.1e-26;  
 Matches 86; Conservative 54; Mismatches 104; Indels 5; Gaps 3;

QY 1 YKDOQKHVGHKACDPEKFGVPELYVDNRRKIRYLOCPCKIYDEBFAELITSH 60  
 DB 55 YRLERRFRFNRH--DQAYIAKGYOPLVNMGEVADVAAYKTKELIAAKTQAISDRINV 111  
 QY 61 HMOQDTLANKKDIYNNHRDLVMAAADICTAITNGEOKGLYVGPFGTKSPILGA 120  
 DB 112 SLPRSYKNSFDDINDVDKLVFKRVADPVEQYPPED-KGLYVGMGICKSTLMDA 170  
 QY 121 IANQLSKK-VRSITIIYLPFIRTLKGFQSGFPEKRLHVRANILMLDDIGAEVTPW 179  
 DB 171 MAHELSEQGAATTLTHPSPFTIDVNAINTGVKKEIDAVKTDADILLDDIGAEQSTGW 230  
 QY 180 VQDEVIGPLHVMHLPFPPSSNDYSELEHHLAMTRDGEKTKARITERYKSLTP 239  
 DB 231 IRDEVIAQVILQYVMELELPFTFTSYNSFKDLBAKLANIKSGDETWQAKRVMERIVTAAKE 290  
 QY 240 YFLSGENR 248  
 DB 291 IHLGENRR 299

## RESULT 11

Q8CWP7 PRELIMINARY; PRT; 298 AA.

AC 08CWP7; 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DE Primosome component (Hellicase loader).  
 GN DNAI OR SPR155.  
 OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 NC Streptococcus.  
 NCBI\_Taxid=171101;

QY 12 KPADCPNFKVGHVPELYVDNRRKIRYLOCPCKIYDEBFAELITSHMOQDTLANK 71  
 DB 64 KKKDSQYIAKGYEPLVNMGEVADVSYLETRILBAQKQAISDRINLVNLPKSYRNIRM 123

RP SEQUENCE FROM N.A.  
 RX MEDLINE=21429245; PubMed=11544234;  
 RA Hoskins J., Albom W.B. Jr., Arnold J., Blazczak L.C., Burgett S.,  
 RA Dehoff B.S., Batem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,  
 RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,  
 RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsumura P.,  
 RA Morris S.M., McHenry M., Mcleaster K., Mundy C.W., Niclas T.I.,  
 RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rokey P.,  
 RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,  
 RA Zook C.A., Baltz R.H., Jaskunas S.R., Kosteck P.R. Jr., Skatrud P.L.,  
 RA Glass J.I.;  
 RT "Genome of the bacterium Streptococcus pneumoniae strain R6."  
 RL J. Bacteriol. 183:5709-5717 (2001).  
 DR EMBL; AE008523; AAL00359.1; -  
 KW Hellicase: Complete proteome.  
 SQ SEQUENCE 298 AA; 34574 MW; 3EA8DBED1DEBFE22 CRC64;

Query Match 28.7%; Score 381; DB 16; Length 298;  
 Best Local Similarity 34.7%; Pred. No. 7e-25;  
 Matches 82; Conservative 50; Mismatches 100; Indels 4; Gaps 3;

QY 15 DCPNFKVGHVPELYVDNRRKIRYLOCPCKIYDEBFAELITSHMOQDTLANKDI 74  
 DB 65 DDDYIAKGYKPYLVNMGEVADVSYLETRILBAKRAIKRLINLPSSLSKNVSFLDV 124  
 QY 75 VNNHRDLVMAAADICTAITNGEOKGLYVGPFGTKSPILGANQLSKK-VRST 133  
 DB 125 YKDDVQRLVLRKMTIEPVNDYRN--NLKGLYVGDVGKSPVVALADLSERKGVST 182  
 QY 134 IYLPFIRTLKGFQSGFPEKRLHVRANILMLDDIGAEVTPWVQDEVIGPLHVM 193  
 DB 183 LHHYSFVIDVNAISDGNVKTLDVDEIKSEVLILLDDIGAEQSTWVNRBEILQVILQYVM 242  
 QY 194 VHELPTFFSSNFDYSELEHHLAMTRDGEKTKARITERYKSLTPYFSGENR 248  
 DB 243 QENLPTFTFTSYNSFDDLEKGFQVKGHGVDETRVEMERIVTAAKE 298

## RESULT 12

Q8B3T8 PRELIMINARY; PRT; 300 AA.

AC 08B3T8; 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DE Hypothetical protein.  
 GN DNAI OR GBS1668.  
 OS Streptococcus agalactiae (serotype III).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 NC Streptococcus.  
 NCBI\_Taxid=216495;

RP SEQUENCE FROM N.A.  
 RC STRAIN=NEM316 / Serotype III;  
 RX MEDLINE=2242508; PubMed=12354221;  
 RA Glaeser P., Ruenick C., Buchrieser C., Chevalier F., Frangoul L.,  
 RA Wsadek T., Zouine M., Couve E., Lalloué L., Poyart C., Titeu-Cuot P.,  
 RA Kunet F.;  
 RT "Genome sequence of Streptococcus agalactiae, a pathogen causing  
 RT invasive neonatal disease."  
 RL Mol. Microbiol. 45:1499-1513 (2002).  
 DR EMBL; AL766852; CAD47327.1; -  
 KW Sagalistic; gbs1668; -  
 SQ SEQUENCE 300 AA; 34679 MW; 9COB4898661661B9 CRC64;

Query Match 27.5%; Score 365; DB 16; Length 300;  
 Best Local Similarity 32.4%; Pred. No. 1.7e-23;  
 Matches 77; Conservative 56; Mismatches 103; Indels 2; Gaps 2;

QY 12 KPADCPNFKVGHVPELYVDNRRKIRYLOCPCKIYDEBFAELITSHMOQDTLANK 71  
 DB 64 KKKDSQYIAKGYEPLVNMGEVADVSYLETRILBAQKQAISDRINLVNLPKSYRNIRM 123

QY 72 KDLYMHRRDLVAMAAADICTAITNGEQVGLYLGPGFGKSFIIIGAIANOLKSKK-V 130  
 DB 124 TDFPINNESKMAKMSQLDPEVETPSPYNNH-KGLYLYGDMGVGKSYLMAAARREISERKGV 182  
 QY 131 RSTIIVLPEFIRTLKGGFKDGSFEKKLHRYREANIIMLDIGAEVTPWVRDEVIGPLH 190  
 DB 183 STTLHFPSPFAIDVKNAISSGTVKDEIDAVKSVPIILIDIGAGQATSWVRDEILQVILQ 242  
 QY 191 YRMVHELPTFPSSNFDVSELEHHLAMTRDGEKTKAARIIEVKSISTPYLSEGNR 248  
 DB 243 HRMLEELPTFPSTNYSFNDLERKMANIKGSDETQAKRVMERVRYALIEPHLEGPNNR 300

## RESULT 13

Q8DY72 PRELIMINARY; PRT; 300 AA.  
 AC Q8DY72;  
 DT 01-MAR-2003 (TReMBLrel. 23, Created)  
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
 DE Primosomal protein DnaI.  
 GN DnaI OR SAG1621.  
 OS Streptococcus agalactiae (serotype V).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 NCBI\_TaxID=216466;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=2603 V/R / Serotype V;  
 RX MEDLINE=22222989; PubMed=12200547;  
 RA Tetteijn H., Maaisgnat V., Cieslewicz M.J., Eissen J.A., Peterson S.,  
 RA Medoff L.C., Wolf A.M., Nelson K.E., Margalit I., Read T.D.,  
 RA Messeri R., Paulsen I.T., Nelson K.E., Brinkac L.M., Daugherty S.C.,  
 RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,  
 RA Radeanu D., Fedotova N.B., Scanlan D., Khoult H., Mulligan S.,  
 RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,  
 RA Jacobini E.T., Bretonni C., Galli G., Mariani M., Vegni P., Malone D.,  
 RA Rinaldo D., Rappuoli R., Telford J.L., Kaper D.L., Grandi G.,  
 RA Fraser C.M.;  
 RT "Complete genome sequence and comparative genomic analysis of an  
 RT emerging human pathogen, serotype V Streptococcus agalactiae".  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).  
 DR EMBL; AB014265; AA00485.1; -.  
 DR TIGR; SAG1621; -.  
 KM Complete proteome.  
 SQ SEQUENCE 300 AA; 34663 MW; 9BED36F86B36B395 CRC64;

Query Match 27.5%; Score 365; DB 16; Length 300;  
 Best Local Similarity 32.4%; Pred. No. 1,7e-23;  
 Matches 77; Conservative 56; Mismatches 103; Indels 2; Gaps 2;

QY 12 KPADCPVGVGHVPELVNMRKIRYLOCPCKIKIDEEFEALITSHMQDITLNAKL 71  
 DB 64 KNDSDOYIAGYEPILVMNEGADVSYLETREIEAQKQAIIDRLNVLPPSYRIRIM 123  
 QY 72 KDLYMHRRDLVAMAAADICTAITNGEQVGLYLGPGFGKSFIIIGAIANOLKSKK-V 130  
 DB 124 TDFPINNESKMAKMSQLDPEVETPSPYNNH-KGLYLYGDMGVGKSYLMAAARREISERKGV 182  
 QY 131 RSTIIVLPEFIRTLKGGFKDGSFEKKLHRYREANIIMLDIGAEVTPWVRDEVIGPLH 190  
 DB 183 STTLHFPSPFAIDVKNAISSGTVKDEIDAVKSVPIILIDIGAGQATSWVRDEILQVILQ 242  
 QY 191 YRMVHELPTFPSSNFDVSELEHHLAMTRDGEKTKAARIIEVKSISTPYLSEGNR 248  
 DB 243 HRMLEELPTFPSTNYSFNDLERKMANIKGSDETQAKRVMERVRYALIEPHLEGPNNR 300

RESULT 14  
 Q97PC7 PRELIMINARY; PRT; 298 AA.  
 ID Q97PC7  
 AC Q97PC7;

DT 01-OCT-2001 (TReMBLrel. 18, Created)  
 DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)  
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
 DE Primosomal protein DnaI.  
 GN Spt711.  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 NCBI\_TaxID=1313;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TIGR4;  
 RX MEDLINE=21357209; PubMed=11463916;  
 RA Tetteijn H., Nelson K.E., Paulsen I.T., Eissen J.A., Read T.D.,  
 RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,  
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radeanu D.,  
 RA Holtzapfel B., Khoult H., Wolf A.M., Utecherback T.R., Hansen C.L.,  
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickenson T., Hickey E.K.,  
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
 RT "Complete genome sequence of a virulent isolate of Streptococcus  
 RT pneumoniae".  
 RL Science 293:498-506(2001).  
 DR EMBL; AB007464; AAK75789.1; -.  
 DR TIGR; SPT711; -.  
 KM Complete proteome.  
 SQ SEQUENCE 298 AA; 34259 MW; C4B5ECAF9C30059 CRC64;

Query Match 27.2%; Score 361; DB 16; Length 298;  
 Best Local Similarity 34.7%; Pred. No. 3,7e-23;  
 Matches 83; Conservative 50; Mismatches 96; Indels 10; Gaps 4;

QY 15 DCNPFKGVGHVPELVNMRKIRYLOCPCKIKIDEEFEALITSHMQDITLNAKL 74  
 DB 65 DTDYIAGKGYKPIVKNHGYADVSEETPELIAEKEAIAIKRLKLNPAISKASIAQV 124  
 QY 75 YNMHRDLV---AMAAADICTAITNGEQVGLYLGPGFGKSFIIIGAIANOLKSKK-V 130  
 DB 125 DLDLRLVFEKTLAFVEQYPAIR-----KGLYLYGDMGVGKSYLMAAARREISERKGV 179  
 QY 131 RSTIIVLPEFIRTLKGGFKDGSFEKKLHRYREANIIMLDIGAEVTPWVRDEVIGPLH 190  
 DB 180 STTLHFPSPFAIDVKNAISSGTVKDEIDAVKSVPIILIDIGAGQATSWVRDEILQVILQ 229  
 QY 191 YRMVHELPTFPSSNFDVSELEHHLAMTRDGEKTKAARIIEVKSISTPYLSEGNR 248  
 DB 240 YRMQENLPTFPSTNYSFNDLERKFAKVGNDETWEARRVERIRYLAETRLLEGVNR 298

## RESULT 15

O9CH9 PRELIMINARY; PRT; 293 AA.  
 ID O9CH9  
 AC O9CH9;  
 DT 01-JUN-2001 (TReMBLrel. 17, Created)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)  
 DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)  
 DE Primosomal protein DnaI.  
 GN DnaI OR L10752.  
 OS Streptococcus lactis (subsp. lactis) (Streptococcus lactis).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.  
 NCBI\_TaxID=1360;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IL1403;  
 RX MEDLINE=21335186; PubMed=11337471;  
 RA Bolotin A., Wincker P., Mauer S., Jallion O., Malarne K.,  
 RA Weisenbach J., Ehrlich S.D., Sorokin A.;  
 RT "The complete genome sequence of the lactic acid bacterium Lactococcus  
 RT lactis ssp. lactis IL1403".  
 RL Genome Res. 11:731-753(2001).  
 DR EMBL; AE006308; AAK04850.1; -.  
 DR InterPro; IPR003593; AAA\_Atpase.

DR SMART; SM00382; AAA; 1.  
KW Complete proteome.  
SQ SEQUENCE 293 AA; 33904 MW; D75C530A69445B3 CRC64;

Query Match 25.8%; Score 342; DB 16; Length 293;  
Best Local Similarity 33.3%; Pred. No. 1.6e-21;  
Matches 85; Conservative 49; Mismatches 97; Indels 24; Gaps 6;

```
QY 2 KQOQKHVDGHRKPADCFNFKGHVPELYVNNRIKIRYLQCPCKIKYDERFPAELITSHH 61
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 57 REHEKEKEKEKRA-----ADGYEPVLIMNHGYADVSY---QTNELAQQAQNLRRMN 108
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 62 ---MORDTLNAKLDIYMAHNRDLDAVMAADDICTAITNGEQVKGXLYGPFGTGKSPIL 118
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 109 IIGLPKDLKQVTLADLDVQRIKPYQALYDF---ITNPKKGGLXLYGDFGVGKSPYL 165
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 119 GAIANQLKSKKQVSTIIYLPETIRTLKGFQDGSPEKK--LHRVREANILMLDDIGAEE 175
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 166 AAMANELAKKGISTILLHYPTFI-----SDLDFDNKAKWVNEIKASQVLVLDIGAEQ 218
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 176 VTPWVDEVIIGPLIHRMVEHLEPTFFSSNFDYSELEHILAMTRDGEKTKAARIIERVKS 235
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 219 NNAWVDSILQVLIQHRMQENLPTFTSNLRMELEQLAETGRADEIWPAKRVMERVKY 278
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 236 LSTPYFLSGENPRNN 250
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 279 LAEMRLBGTNRHD 293
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: December 15, 2003, 15:15:58  
Job time : 29.8858 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 15:11:40 ; Search time 45.6369 Seconds  
(without alignments)  
1088.625 Million cell updates/sec

Title: US-09-689-952-2  
Perfect score: 1643  
Sequence: 1 MCGGSGIMKQFSLINTSD.....ERVKSLSTPYPLSGENFRNN 313

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues  
Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A.Geneseq.19Jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1643	100.0	313	22	AA847317
2	1378	83.9	307	22	ABP38947
3	1353	82.3	299	22	AA682049
4	1353	82.3	299	22	AA682337
5	1328	80.8	250	22	AA847316
6	851	51.8	164	22	AA847315
7	617.5	37.6	155	22	AA681811
8	552.5	33.6	307	23	AB849355
9	429.5	26.1	300	23	ABP27829

10	403.5	24.6	293	23	AB854065
11	393	23.9	92	23	ABP09981
12	393	23.9	300	23	ABP27828
13	392.5	23.9	298	24	ABU02187
14	112	6.8	291	24	AB853757
15	110	6.7	866	22	AAU38917
16	110	6.7	872	20	AAV34745
17	109.5	6.7	294	23	AB854759
18	109	6.6	630	23	AB848390
19	108.5	6.6	400	21	AA634347
20	107	6.5	331	23	ABP65341
21	107	6.5	331	23	ABP65354
22	107	6.5	331	23	ABP65347
23	106.5	6.5	797	21	AA181879
24	106.5	6.5	897	22	AB862660
25	106	6.5	435	21	AA634039
26	106	6.5	500	21	AA634038
27	105	6.4	248	22	AA698965
28	105	6.4	479	22	ABG18303
29	105	6.4	798	22	AA696715
30	105	6.4	1886	19	AAW54241
31	104.5	6.4	386	22	ABG25088
32	104	6.3	399	21	AA626569
33	103.5	6.3	397	22	AB859915
34	103.5	6.3	399	21	AA646306
35	103	6.3	308	24	ABU01560
36	102.5	6.2	304	23	ABP40308
37	102	6.2	389	24	ABJ25711
38	102	6.2	389	24	ABJ26311
39	102	6.2	453	23	ABP30561
40	102	6.2	484	23	ABP27967
41	102	6.2	1935	23	ABG31649
42	101.5	6.2	784	22	ABG30180
43	101.5	6.2	827	23	ABP73784
44	101	6.1	451	24	ABJ26363
45	101	6.1	2539	21	AA181898

## ALIGNMENTS

RESULT 1	AA847317	standard; Protein; 313 AA.
XX	AA847317;	
XX	29-AUG-2001 (first entry)	
DE	S. aureus DnaI.	
XX	DnaI; S. aureus; inhibitor; bacteriophage 77; ORF 104; phage 77ORF104;	
KW	screening assay.	
XX	Staphylococcus aureus.	
OS	Staphylococcus aureus.	
PN	WO200146383-A2.	
PD	28-JUN-2001.	
XX	21-DEC-2000; 2000WO-US35180.	
PF	22-DEC-1999; 99US-0470512.	
PR	12-OCT-2000; 2000US-0689952.	
XX	(PHAG-) PHAGETECH INC.	
PA	(WILLI) WILLIAMS K M.	
XX	Pelletier J, Gros P, Dubow M;	
XX	WPI; 2001-418052/44.	
DR	N-PSDB; AAC86105.	
XX		

Lactococcus lactis  
Human ORF protein  
Streptococcus poly  
S. pneumoniae type  
Lactococcus lactis  
C. pneumoniae CTII  
C. pneumoniae prot  
Lactococcus lactis  
Lactaria monocytog  
Zea mays protein f  
Bifidobacterium lo  
Bifidobacterium lo  
Plasmodium falcipar  
Drosophila melanog  
Zea mays protein f  
Zea mays protein f  
E. coli growth and  
Novel human diagno  
Putative P. abyssi  
Rattus norvegicus  
Novel human diagno  
Arabidopsis thaliana  
Drosophila melanog  
Arabidopsis thaliana  
S. pneumoniae type  
Staphylococcus epi  
Aspergillus fumig  
Aspergillus fumig  
Streptococcus poly  
Streptococcus poly  
Amino acid distrib  
Novel human diagno  
Candida albicans e  
Aspergillus fumig  
Plasmodium falcipar

PT Novel DnaI polypeptides useful for treating and diagnosing microbial,  
PT preferably bacterial, diseases such as those caused by *Staphylococcus*  
XX aureus -  
XX  
XX Disclosure, Fig 1, 107bp, English.  
XX  
XX This sequence shows DnaI derived from *S. aureus*. *S. aureus* DnaI  
CC interacts with a growth-inhibitory bacteriophage 77 ORF 104  
CC gene product, to form the basis of a screening assay. DnaI  
CC polypeptides and polynucleotides are useful for treating microbial,  
CC preferably bacterial, especially *Staphylococcus*, infections. DnaI  
CC polypeptides and polynucleotides are useful for biological, diagnostic,  
CC prophylactic, clinical and therapeutic use, and as components in  
CC databases useful for search analyses as well as in sequence analysis  
CC algorithms.  
XX  
XX Sequence 313 AA;  
SQ

Query Match Best Local Similarity 100.0%; Score 1643; DB 22; Length 313;

Matches 313; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGSGQSIMKQFKSIINTSDPEKRIEIKKEVINDDPVKQFLERHRELTMAMIDEDLVN 60  
DB 1 MGSGQSIMKQFKSIINTSDPEKRIEIKKEVINDDPVKQFLERHRELTMAMIDEDLVN 60  
QY 61 LQRYKDDQKHYDGHKPADCPNFKVGHVPELYVNNRIRIKRYLQCPCKIKYDEERPEAEI 120  
DB 61 LQRYKDDQKHYDGHKPADCPNFKVGHVPELYVNNRIRIKRYLQCPCKIKYDEERPEAEI 120  
QY 121 TSHHMQRDTLANAKLDIYNNHRDLVMAADICTAITNGEQVKGILYGFPGTGSFTI 180  
DB 121 TSHHMQRDTLANAKLDIYNNHRDLVMAADICTAITNGEQVKGILYGFPGTGSFTI 180  
QY 181 LGAIANOLSKKQVRSITIIYLPFIRITLKGFGDGSFEKKLHVRVREANITMLDIGAEEVT 240  
DB 181 LGAIANOLSKKQVRSITIIYLPFIRITLKGFGDGSFEKKLHVRVREANITMLDIGAEEVT 240  
QY 241 PWRDEVIGPLHYRMVHELPTFSSNFDYSELEHNLAMTRDGEKTKARIIERYVLS 300  
DB 241 PWRDEVIGPLHYRMVHELPTFSSNFDYSELEHNLAMTRDGEKTKARIIERYVLS 300  
QY 301 TPYFLSGENFRNN 313  
DB 301 TPYFLSGENFRNN 313

## RESULT 2

ABP38947

ID ABP38947 standard; Protein; 307 AA.

XX ABP38947;

XX 24-JUL-2002 (first entry)

XX Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3792.

XX Staphylococcus epidermidis, open reading frame; ORF; bacterial infection;  
XX antibacterial; gene therapy.

XX Staphylococcus epidermidis.

XX US6380370-B1.

XX 30-APR-2002.

XX 13-AUG-1998; 98US-0134001.

XX 14-AUG-1997; 97US-055779P.

XX 08-NOV-1997; 97US-064964P.

XX (GENO-) GENOME THERAPEUTICS CORP.  
XX

PI Doucette-Stamm LA, Bush D;  
XX WPI: 2002-38125/41.  
DR N-PSDB; ABN91492.  
XX  
XX Novel isolated nucleic acid encoding a *Staphylococcus epidermidis*  
PT polypeptide, useful for diagnosing and treating bacterial infections -  
XX  
XX Disclosure; SEQ ID 3792; 267bp; English.  
PS

CC ABN90538 to ABN93374 represent *Staphylococcus epidermidis* open reading  
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
CC given in ABP35124 to ABP37960. The *S. epidermidis* sequences have  
CC antibacterial activity and can be used in gene therapy. The sequences  
CC can also be used in the diagnosis and treatment of bacterial infections,  
CC particularly *S. epidermidis* infections. The sequences can be used to  
CC screen for compounds able to interfere with the *S. epidermidis* life  
CC cycle or inhibit *S. epidermidis* infection.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC USPTO web site.

SQ Sequence 307 AA;

Query Match Best Local Similarity 83.9%; Score 1378; DB 23; Length 307;

Matches 254; Conservative 34; Mismatches 19; Indels 0; Gaps 0;

QY 7 IMKQPSIINTSDPEKRIEIKKEVINDDPVKQFLERHRELTMAMIDEDLVNLOEKYD 66  
DB 1 LKMSFNNIMSDSONLDRIOIKIKONVINDDVGHFLEKRSNITNEMIDEDLVNLOEKYD 60  
QY 67 QQGHVGHKPADCPNFKVGHVPELYVNNRIRIKRYLQCPCKIKYDEERPEAEIITSHMQ 126  
DB 67 QQGHVGHKPADCPNFKVGHVPELYVNNRIRIKRYLQCPCKIKYDEERPEAEIITSHMQ 120  
QY 127 RDTLANAKLDIYNNHRDLVMAADICTAITNGEQVKGILYGFPGTGSFTIGAIAN 186  
DB 127 RDTLANAKLDIYNNHRDLVMAADICTAITNGEQVKGILYGFPGTGSFTIGAIAN 180  
QY 187 QLSKQVRSITIIYLPFIRITLKGFGDGSFEKKLHVRVREANITMLDIGAEEVT 246  
DB 187 QLSKQVRSITIIYLPFIRITLKGFGDGSFEKKLHVRVREANITMLDIGAEEVT 240  
QY 247 VIGPLHYRMVHELPTFSSNFDYSELEHNLAMTRDGEKTKARIIERYVLS 306  
DB 247 VIGPLHYRMVHELPTFSSNFDYSELEHNLAMTRDGEKTKARIIERYVLS 300  
QY 307 GENFRNN 313  
DB 307 GENFRNN 307

## RESULT 3

AAG82049

ID AAG82049 standard; Protein; 299 AA.

XX AAG82049;

XX 03-SEP-2001 (first entry)

XX *S. epidermidis* open reading frame protein sequence SEQ ID NO:1192.XX Staphylococcus epidermidis SRI strain; infection; diagnosis;  
XX vaccination; endocarditis.

XX Staphylococcus epidermidis.

XX WO200134809-A2.

XX 17-MAY-2001.

XX 09-NOV-2000; 2000WO-US30782.  
XX

XX		09-NOV-1999;	99US-0164258.
PR	FA	(GLAX ) GLAXO GROUP LTD.	
XX	PI	Kimmerly WJ;	
XX	DR	WPI: 2001-316495/33.	
DR	N-PSDB; AAH52899.		
PT		Nucleic acids encoding polypeptides from Staphylococcus epidermidis,	
XX		useful for vaccinating against infections, e.g. endocarditis -	
PS		Claim 18, Page 342; 2188bp; English.	
XX		AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides	
CC		(II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.	
CC		(I) and (II) can have antibacterial activity and therefore can be used	
CC		in vaccination. The nucleic acids (I) may be used to produce the	
CC		S. epidermidis polypeptides (II) via the production of vectors	
CC		containing them which are used to produce hosts cells which express the	
CC		polypeptides. The polypeptides (II) (and/or nucleic acids) may then be	
CC		used to vaccinate subjects and to raise antibodies against the bacteria.	
CC		The polypeptides may also be used to assay for other inhibitors of their	
CC		activity and therefore identify compounds that may be used for the	
CC		treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to	
CC		AAH55090 represent specifically claimed S. epidermidis genomic DNA	
CC		polynucleotide sequences from the present invention. AAH55091 to	
CC		AAH55098 represent oligonucleotide sequences and primers which are used	
CC		in the exemplification of the present invention.	
CC		N.B. The present invention specifically claims all the polynucleotide	
CC		sequences given in the sequence listing of the present specification,	
CC		however the sequence listing only goes up to SEQ ID NO:4454 so even	
CC		though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,	
CC		no sequences are present for SEQ ID NO:4455 to 4464.	
SO		Sequence 299 AA:	
QY		Query Match 82.3%; Score 1353; DB 22; Length 299;	
Db		Best Local Similarity 84.5%; Pred. No. 2.7e-113;	
		Matches 250; Conservative 30; Mismatches 16; Indels 0; Gaps 0.	
QY		18 SQDEPKRIEKKKEIVNDPVKQFLFARPAELTNAMIDEDLVLOFYDQQKHNYGAKPA 77	
Db		4 SQNDLKRIOIKQNIYINDTVGVGFPEKRSRNTINMIEDIALNVLOEYDQQKVYGYGHRYD 63	
QY		78 DCPNFKVSHVELYYDNRRIRIKRYIQCPCKIKYDEREPREALITSHHQRDTLNKLMDI 137	
Db		64 DCPNFVKSHVPELYENERIKIRYPCEPKIKHDERPDSQSILTISHHQRTDLNLAKLDI 123	
QY		138 YANHRRRLDVMAAADICTAATNGEQVKGVLVYGFPGFGKSFETLGAINOLSKSKVRSTI 197	
Db		124 YNNRRERLDVMAADQICTAITNDKRVKGLVYGFPGFGKSIYLGAINGOLSKSQISSTI 183	
QY		198 IYLPFEFITLKGGFDGSPFEKKLHVREANITIMDDIGAEVTPVNRDEVIGPLIHYRMV 257	
Db		184 VYLPEFITLKGGFDGSGFEKKLQVRERANITIMDDIGAEVTPVNRDEVIGPLIHYRMV 243	
QY		258 HELPPFSNSNDYSLEHHLMATRGGEKTKAARIIEVKSISTPYFLSGENFRNN 313	
Db		244 HELPFERSNFYSLEHLSTTRGTGTEKTKAARIIEKITLSTPYFLTGKNFRNN 299	
RESULT 4			
AAAG82337			
ID	AAAG82337	standard; Protein; 299 AA.	
XX	AAAG82337;		
AC			
DT	03-SEP-2001	(first entry)	
XX			
DE	S. epidermidis open reading frame protein sequence SEQ ID NO:11768.		
XX			

[illegible]

RESULT 5  
AAB47316  
ID AAB47316 standard; Protein; 250 AA.  
XX  
AC AAB47316;  
XX  
DT 29-AUG-2001 (first entry)  
XX  
DE Amino acids 64-313 of S. aureus DnaI.  
XX  
KM DnaI; S. aureus; inhibitor; bacteriophage 77; ORF 104; phage 77ORF104;  
XX screening assay.  
XX  
OS Staphylococcus aureus.  
XX  
PN MO200146383-A2.  
XX  
PD 28-JUN-2001.  
XX  
PF 21-DEC-2000; 2000MO-US35180.  
XX  
PR 22-DEC-1999; 99US-0470512.  
XX 12-OCT-2000; 2000US-0689952.  
XX  
PA (PHAG-) PHAGETECH INC.  
XX (WILL/) WILLIAMS K M.  
XX  
PI Pelletier J, Gros P, Dubow M;  
XX WPI; 2001-418052/44.  
XX  
DR Novel DnaI polypeptides useful for treating and diagnosing microbial,  
XX preferably bacterial, diseases such as those caused by Staphylococcus  
PT aureus -  
XX  
PS Disclosure; Fig 15; 107pp; English.  
XX  
CC This sequence shows a DnaI polypeptide derived from S. aureus. S.  
CC aureus DnaI interacts with a growth-inhibitory bacteriophage 77 ORF  
CC 104 gene product, to form the basis of a screening assay. DnaI  
CC polypeptides and polynucleotides are useful for treating microbial,  
CC preferably bacterial, especially Staphylococcus, infections. DnaI  
CC polypeptides and polynucleotides are useful for biological, diagnostic,  
CC prophylactic, clinical and therapeutic use, and as components in  
CC databases useful for search analyses as well as in sequence analysis  
CC algorithms.  
XX  
SQ Sequence 250 AA;  
XX  
Query Match 80.8%; Score 1328; DB 22; Length 250;  
Best Local Similarity 100.0%; Pred. No. 3.7e-111;  
Matches 250; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 64 YKQOQKHGDKFPADCFNFKGVHPELVNDRIRIKIRYLOCPCKIKYDEERFEAEILITSH 123  
DB 1 YKQOQKHGDKFPADCFNFKGVHPELVNDRIRIKIRYLOCPCKIKYDEERFEAEILITSH 60  
XX  
QY 124 HMQRDITLNAKLKDIYNNHRRRLDVAMADDICTAITNGEOVKGLYLYGPGTGKSFILGA 183  
DB 61 HMQRDITLNAKLKDIYNNHRRRLDVAMADDICTAITNGEOVKGLYLYGPGTGKSFILGA 120  
XX  
QY 184 IANOLSKKVRSTIYLPFIRTLKGGFKDGSFEKTLHVRREANILMLDIDIGAEVTPMW 243  
DB 121 IANOLSKKVRSTIYLPFIRTLKGGFKDGSFEKTLHVRREANILMLDIDIGAEVTPMW 180  
XX  
QY 244 RDEVIGPLHYRWVHELPTFFSSNFDYSELEHHLAMTRDGEKTKAARIIRVKSISTPY 303  
DB 181 RDEVIGPLHYRWVHELPTFFSSNFDYSELEHHLAMTRDGEKTKAARIIRVKSISTPY 240  
XX  
QY 304 FLSGENFRNN 313  
DB 241 FLSGENFRNN 250

RESULT 6  
AAB47315  
ID AAB47315 standard; Protein; 164 AA.  
XX  
AC AAB47315;  
XX  
DT 29-AUG-2001 (first entry)  
XX  
DE Amino acids 150-313 of S. aureus DnaI.  
XX  
KM DnaI; S. aureus; inhibitor; bacteriophage 77; ORF 104; phage 77ORF104;  
XX screening assay.  
XX  
OS Staphylococcus aureus.  
XX  
PN MO200146383-A2.  
XX  
PD 28-JUN-2001.  
XX  
PF 21-DEC-2000; 2000MO-US35180.  
XX  
PR 22-DEC-1999; 99US-0470512.  
XX 12-OCT-2000; 2000US-0689952.  
XX  
PA (PHAG-) PHAGETECH INC.  
XX (WILL/) WILLIAMS K M.  
XX  
PI Pelletier J, Gros P, Dubow M;  
XX WPI; 2001-418052/44.  
XX  
DR N-PSDB; AAC6104.  
XX  
DR Novel DnaI polypeptides useful for treating and diagnosing microbial,  
XX preferably bacterial, diseases such as those caused by Staphylococcus  
PT aureus -  
XX  
PS Claim 41; Fig 15; 107pp; English.  
XX  
CC This sequence shows a DnaI polypeptide derived from S. aureus. S.  
CC aureus DnaI interacts with a growth-inhibitory bacteriophage 77 ORF  
CC 104 gene product, to form the basis of a screening assay. DnaI  
CC polypeptides and polynucleotides are useful for treating microbial,  
CC preferably bacterial, especially Staphylococcus, infections. DnaI  
CC polypeptides and polynucleotides are useful for biological, diagnostic,  
CC prophylactic, clinical and therapeutic use, and as components in  
CC databases useful for search analyses as well as in sequence analysis  
CC algorithms.  
XX  
SQ Sequence 164 AA;  
XX  
Query Match 51.8%; Score 851; DB 22; Length 164;  
Best Local Similarity 100.0%; Pred. No. 1.6e-68;  
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 150 AADDICTAITNGEOVKGLYLYGPGTGKSFILGAIANOLSKKVRSTIYLPFIRTLK 209  
DB 1 AADDICTAITNGEOVKGLYLYGPGTGKSFILGAIANOLSKKVRSTIYLPFIRTLK 60  
XX  
QY 210 GFPOGSFEKTLHVRREANILMLDIDIGAEVTPWVRDEVIGPLHYRWVHELPTFFSSNFD 269  
DB 61 GFPOGSFEKTLHVRREANILMLDIDIGAEVTPWVRDEVIGPLHYRWVHELPTFFSSNFD 120  
XX  
QY 270 YSELEHHLAMTRDGEKTKAARIIRVKSISTPYFLSGENFRNN 313  
DB 121 YSELEHHLAMTRDGEKTKAARIIRVKSISTPYFLSGENFRNN 164  
XX  
RESULT 7  
AAG81811  
ID AAG81811 standard; Protein; 155 AA.  
XX  
AC AAG81811;

XX 03-SEP-2001 (first entry)  
 DT S. epidermidis open reading frame protein sequence SEQ ID NO:716.  
 DE Staphylococcus epidermidis SRI strain; infection; diagnosis;  
 KW vaccination; endocarditis.  
 XX Staphylococcus epidermidis.  
 OS WO200134809-A2.  
 PN 17-MAY-2001.  
 PD 09-NOV-2000; 2000MO-US30782.  
 PF 09-NOV-1999; 99US-0164258.  
 PR (GLAXO) GLAXO GROUP LTD.  
 PA Kimmberly WJ;  
 PI WPI; 2001-316495/33.  
 DR N-PSDB; AAH52661.  
 XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
 PT useful for vaccinating against infections, e.g. endocarditis -  
 PS Claim 18; Page 223; 2188bp; English.  
 XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
 CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis  
 CC (I) and (II) can have antibacterial activity and therefore can be used  
 CC in vaccination. The nucleic acids (I) may be used to produce the  
 CC S. epidermidis polypeptides (II) via the production of vectors  
 CC containing them which are used to produce hosts cells which express the  
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
 CC used to vaccinate subjects and to raise antibodies against the bacteria.  
 CC The polypeptides may also be used to assay for other inhibitors of their  
 CC activity and therefore identify compounds that may be used for the  
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
 CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
 CC polynucleotide sequences from the present invention. AAH55091 to  
 CC AAH55098 represent oligonucleotide sequences and primers which are used  
 CC in the exemplification of the present invention.  
 CC N.B. The present invention specifically claims all the polynucleotide  
 CC sequences given in the sequence listing of the present specification,  
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
 CC no sequences are present for SEQ ID NO:4455 to 4464.  
 XX  
 SQ Sequence 155 AA;  
 Query Match 37.6%; Score 617.5; DB 22; Length 155;  
 Best Local Similarity 77.0%; Pred. No. 1,4e-47;  
 Matches 114; Conservative 19; Mismatches 12; Indels 3; Gaps 1;  
 QY 18 SQDFEKRIEIKKEVINDDPVKQFLFAHRAELTNAMIDEDLVLYQEKYQOKHYDGKFA 77  
 DB 4 SQMLDKRIQIKQNVINDTVKAFLEKRNRSNITNEMIDEDLVLYQEKYQOKHYDGKFA 63  
 QY 78 DCNPFKGVHPELVYDNNRKIRYLQCPCKIKYDERFEALITSHHMQDTLNAKLKDI 137  
 DB 64 DCNPFKGVHPELVYDNNRKIRYLQCPCKIKYDERFEALITSHHMQDTLNAKLKDI 123  
 QY 138 YNNHRDLVAVMAADICTAITNGEYK 165  
 DB 124 YNNHRDLVAVMAADICTAITNGEYK 148  
 RESULT 8  
 ABB49355  
 ID ABB49355 standard; Protein; 307 AA.

XX ABB49355;  
 AC 05-FEB-2002 (first entry)  
 DT Listeria monocytogenes protein #2059.  
 DE Listeria monocytogenes protein #2059.  
 KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 KM vitamin B12; bacterial infection; disease.  
 XX Listeria monocytogenes.  
 OS WO200177335-A2.  
 PN 18-OCT-2001.  
 PD 11-APR-2001; 2001WO-FR01118.  
 PF 11-APR-2000; 2000FR-0004629.  
 PR (INSP) INSP PASTEUR.  
 PA Buchrieser C, Frangeul L, Couve E, Ruenigk C, Faini H, Dehoux P,  
 PI Dussurgeat O, Chetoui F, Nedjati H, Glaser P, Kunet F, Cossart P,  
 PI Daniels J, Goebel W, Krefz J, Kuhn M, Ng E, Vazquez-Boland JA,  
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A,  
 PI Chakraborty T, Domann E, Hain T, Berche P, Chardit A, Durant L,  
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N,  
 PI Madueno E, De Pablos B, Weiland J, Kaerst U, Entian K, Hauf J;  
 PI Rose M, Voss H;  
 XX WPI; 2002-010914/01.  
 DR Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
 XX and prevention of Listeria and related bacterial infections, and  
 PT related polypeptides -  
 PS Claim 6; SEQ ID NO 2060; 192bp; French.  
 XX The present invention relates to the genome sequence of Listeria  
 CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of  
 CC it are useful for selecting probes and primers for detecting genes in L.  
 CC monocytogenes and related organisms, and for studying genetic  
 CC polymorphisms and other genomes. The present sequence is a protein  
 CC encoded by the genome sequence of the present invention. Proteins  
 CC expressed from the genome sequence are useful for raising specific  
 CC antibodies, identification of L. monocytogenes and related organisms, and  
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
 CC B12. The genome sequence and proteins encoded by it are also useful for  
 CC selecting compounds that regulate gene expression and cell replication  
 CC and modulate L. monocytogenes-related diseases. In addition, the genome  
 CC sequence and proteins encoded by it are useful in pharmaceutical and  
 CC vaccines compositions for the treatment or prevention of infections by L.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 307 AA;  
 Query Match 33.6%; Score 552.5; DB 23; Length 307;  
 Best Local Similarity 37.8%; Pred. No. 2,4e-41;  
 Matches 113; Conservative 69; Mismatches 104; Indels 13; Gaps 4;  
 QY 19 ODFFEKRIEIKKEVINDDPVKQFLFAHRAELTNAMIDEDLVLYQEKYQOKHYDGKFA 78  
 DB 15 RDBFEKRYQGLKQVLYHYPIDQFFKHKKEVTEQVNLQSLUYEMTGHKKFTGE 71  
 QY 79 CENPFKGVHPELVYDNNRKIRYLQCPCKIKYDERFEALITSHHMQDTLNAKLKDI 138  
 DB 72 -ETLMPGVAPKLVNLGEFTVYTYPTKEKIEDKRAVRRIRSLYMPKQVVDANLADPY 130  
 QY 139 NNNHRDLVAVMAADICTAITN-----GRQVGLVLYGFTGSKPILGAIANNQSKKV 193

Db 131 TDESR---QALVEAVQFLNNYPPKSGERYKGLFHIGSFETGKSYLLGALAKELAKGI 187  
 QY 194 RSTIIYPERFIRLKGKFGKSGPEKRLHRYEANIIMLDIGAEVTPMWRDEVITGPIH 253  
 Db 188 STIIVYPERFIRKVGKISIDVTGEKIQPAKETVIMLDIGAEVTPMWRDEVITGPIH 247  
 QY 254 YRMVHELPTFPSSNFDYSELEHILAMTRDG-EETKARILIERVKSISTPYFLSGENP 311  
 248 FRMGELPTFPSSNFDYSELEHILAMTRDG-EETKARILIERVKSISTPYFLSGENP 306

RESULT 9  
 ABP27829  
 ID ABP27829 standard; Protein; 300 AA.

AC ABP27829;  
 DT 02-JUL-2002 (first entry)  
 DE Streptococcus polypeptide SEQ ID NO 4834.  
 XX  
 KM Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 KM group A streptococcus; Streptococcus pyogenes; antibacterial;  
 KM antiinflammatory; infection; vaccine; meningitis; gene therapy.  
 XX  
 OS Streptococcus pyogenes.  
 XX  
 PN W0200234771-A2.  
 XX  
 PD 02-MAY-2002.  
 XX  
 PF 29-OCT-2001; 2001WO-GB04789.  
 XX  
 PR 27-OCT-2000; 2000GB-0026333.  
 PR 24-NOV-2000; 2000GB-0028727.  
 PR 07-MAR-2001; 2001GB-0005640.  
 XX  
 PA (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Telford J, Maignani V, Margarit Ros YI, Grandi G, Fraser C;  
 PI Tettelin H;  
 XX  
 DR WPI; 2002-352536/38.  
 DR N-PSDB; ABN68460.  
 XX  
 PT New Streptococcus protein for the treatment or prevention of infection  
 PT or disease caused by Streptococcus bacteria, such as meningitis, and  
 PT for detecting a compound that binds to the protein -  
 XX  
 PS Claim 1; Page 3645; 4525pp; English.

The invention relates to a protein (ABP25413-ABP30895) from group B  
 Streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS  
 (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 the specification. The proteins have antibacterial and antiinflammatory  
 activity. (1), nucleic acids encoding (1), ABN6044-ABN71526 and  
 antibodies that bind (1) are used in the manufacture of medicaments for  
 the treatment or prevention of infection or disease caused by  
 Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (1) are used to detect Streptococcus in a  
 CC biological sample. (1) is used to determine whether a compound binds to  
 CC (1). A composition comprising (1) or a nucleic acid encoding (1), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (1) may be used to recombinantly produce (1) and may be  
 CC used in gene therapy. Antibodies to (1) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins.  
 XX  
 XX Sequence 300 AA;

Query Match 26.14; Score 429.5; DB 23; Length 300;  
 Best Local Similarity 31.24; Pred. No. 2,7e-30;  
 Matches 94; Conservative 74; Mismatches 126; Indels 7; Gaps 4;

QY 12 KSIIINSQDEFEKIKKEKVINDDPKQFLHARAEITAMIDEDLVNLOEYKDOCKY 71  
 Db 6 ETWAKLGQONRVNSDOIQTILADPEVASPISQH-HLSGQINTLSKRNQFLVERQKY 63  
 QY 72 DGKHFADCFNFAVGHVELVYDNNRIKIRYLQCPCKIKYEEFPAELTSHMQRDITLN 131  
 Db 64 ---QKDPSTIAGKYQDILMAGYADVSTLERKEVLAQKQAISRILQVSLPKSYRH 120  
 QY 132 AKLDIYMNHRDLVDVMAADDICTAITNGEYKGLVLYGPFQSGSPILGAIAMQSK 191  
 Db 121 IHLSDIDVNNASHWEAFSAIILDFVEQYPSAEQ-KGLYLYGDMQIGKSYLLAAAHLESEK 179  
 QY 192 K-VRSITIIYPERFIRLKGKFGKSGPEKRLHRYEANIIMLDIGAEVTPMWRDEVITGP 250  
 Db 180 KGVSTIILHPPSPFAIDKNAISNGSVKEIDAVKPVLLILDIGAEQATSWRDEVITQV 239  
 QY 251 LHYRMVHELPTFPSSNFDYSELEHILAMTRDGEETKARILIERVKSISTPYFLSGENP 310  
 Db 240 ILQYRMHELPTFPSTNYSFADLERKWAITKSGDETQAKRVVERVYILAREFHLEGANR 299

QY 311 R 311  
 Db 300 R 300

RESULT 10  
 ABB54065  
 ID ABB54065 standard; Protein; 293 AA.

AC ABB54065;  
 DT 16-MAY-2002 (first entry)  
 DE Lactococcus lactis protein dnaI.  
 XX  
 KM Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.  
 XX  
 OS Lactococcus lactis IL1403.  
 XX  
 PN FR2807446-A1.  
 XX  
 PD 12-OCT-2001.  
 XX  
 PF 11-APR-2000; 2000FR-0004630.  
 PF 11-APR-2000; 2000FR-0004630.  
 PR 11-APR-2000; 2000FR-0004630.  
 XX  
 PA (INRG) INRA INST NAT RECH AGRONOMICQUE.  
 XX  
 PI Bolotine A, Sorokine A, Renault P, Ehrlich SD;  
 PI WPI; 2002-043418/06.  
 XX  
 PT New nucleotide sequence useful in the identification or Lactococcus  
 PT lactis and related species -  
 XX  
 PS Claim 6; SEQ ID No 767; 2504pp; French.

The present invention is related to a Lactococcus lactis nucleotide  
 CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The  
 CC nucleic acid sequence is useful in the detection and/or amplification of  
 CC related species. The proteins of the invention are useful for the  
 CC biosynthesis or biodegradation of a composition of interest. The  
 CC invention helps research in lactic bacteria, particularly useful in the  
 CC production of yogurt and cheese.  
 CC Note: The sequence data for this patent is based on equivalent patent  
 CC W020017734 (published 18-Oct-2001) which is available in electronic  
 CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Disclosure; SEQ ID 19944; 1037pp; English.

The present invention describes substantially purified human proteins (referred to as open reading frame, ORFX, where X is 1-1149) (see Table 1 in the specification). ABRN15762 to ABRN27252 encode the human ORFX proteins given in ABRN00010 to ABRN15500. ORFX proteins are useful for treating or preventing a pathology associated with an ORFX-associated disorder in humans, and in the manufacture of a medicament for treating a syndrome associated with ORFX-associated disorder. ORFX polynucleotide sequences can be used in gene therapy. ORFX sequences can be used in the treatment of cancer, hyperproliferative disorders, cirrhosis of liver, psoriasis, benign tumors, keloid, degenerative disorders, haemorrhage, osteoarthritis, neurodegenerative disorders, disorders related to organ transplantation, cardiovascular diseases, diabetes mellitus, systemic lupus erythematosus, hypertension, hypothyroidism, cholesterol ester storage disease, various immune deficiencies and disorders, infectious diseases, autoimmune disorders such as multiple sclerosis, rheumatoid arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. ORFX proteins are also useful for treating burns, incisions, ulcers, for treating osteoporosis, bone degenerative disorders, and periodontal disease, and for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues and conditions resulting from systemic cytokine damage.

N.B. The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

Sequence 92 AA:

Query Match 23.9%; Score 393; DB 23; Length 92;  
 Best Local Similarity 81.3%; Pred. No. 1.1e-27;  
 Matches 74; Conservative 10; Mismatches 7; Indels 0; Gaps 0;

163 QVKKLYLYGPGTSGKSFILGAIANOLKSKYRSTIYLPFIRTKGFGDGSFEKTLHR 222  
 2 RVKGLYIHGFGTSGKSFILKSIANOLKSKVPSITVLPYIRLTKGFGDGSYNELKLR 61

223 VREANIIMDDIGAEVYTPWVDEVIGPLIH 253  
 62 IREANIIMDDIGAEVDTWVRDEVIGPLIH 92

RESULT 12  
 ABR27828  
 ABR27828 standard; Protein; 300 AA.  
 ABR27828;  
 02-JUL-2002 (first entry)  
 Streptococcus polypeptide SEQ ID NO 4832.  
 Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 group A streptococcus; Streptococcus pyogenes; antibacterial;  
 antiinflammatory; infection; vaccine; meningitis; gene therapy.  
 Streptococcus agalactiae.  
 WO200234771-A2.  
 02-MAY-2002.  
 29-OCT-2001; 2001WO-GB04789.  
 27-OCT-2000; 2000GB-0026333.  
 24-NOV-2000; 2000GB-0028727.  
 07-MAR-2001; 2001GB-0005640.  
 (CHIR-) CHIRON SPA.  
 (GENO-) INST GENOMIC RES.  
 Telford J, Maignani V, Margarit Ros YI, Grandi G, Frazer C;

PI Tectelin H;  
 XX WPI: 2002-352536/38.  
 DR N-PSDB; ABN6459.  
 XX  
 PT New Streptococcus protein for the treatment or prevention of infection  
 PT or disease caused by Streptococcus bacteria, such as meningitis, and  
 PT for detecting a compound that binds to the protein -  
 PS  
 PS Claim 1; Page 3645; 4525pp; English.  
 XX  
 XX The invention relates to a protein (ABP25413-ABP30895) from group B  
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 CC the specification. The proteins have antibacterial and antiinflammatory  
 CC activity. (1), nucleic acids encoding (1), ABN66044-ABN71526 and  
 CC antibodies that bind (1) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (1) are used to detect Streptococcus in a  
 CC biological sample. (1) is used to determine whether a compound binds to  
 CC (1). A composition comprising (1) or a nucleic acid encoding (1), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (1) may be used to recombinantly produce (1) and may be  
 CC used in gene therapy. Antibodies to (1) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins.  
 XX  
 SQ Sequence 300 AA;  
 Query Match 23.9%; Score 393; DB 23; Length 300;  
 Best Local Similarity 28.5%; Pred. No. 5, 1e-27;  
 Matches 88; Conservative 80; Mismatches 127; Indels 14; Gaps 5;  
 QY 4 GGSIMQFKSIINTSGPEKRIEIKKEVINDPVKQPLEAHRALTNAMIDEDLVLOE 63  
 DB 5 GQALENGGRVBRNTN-----DELQIMILADQVAFIKTH--QLSGREINISMSKFNQ 55  
 QY 64 YKQKQKHYDGHKFPDCNPFVKGHVELVYNNRKIKIYLCQPCCKIKYDEREFAELTSH 123  
 DB 56 FLIERQKPF---KNKDSQYIAKGYEPILVMEGVADSVLTRELIIEKQKQALSDRLNLV 112  
 QY 124 HMORDTLNAKLKDLYNNHRDLVDVMAADICTAITNGEYQKGLYLGPFGTGKSFILGA 183  
 DB 113 NLPKSYNNIMTPDINNBSRMKMSQLDPVEYTPYNN-KGLYLYXGDMGVKSYLMAA 171  
 QY 184 IANQLSKK-VRSITTYLPEFIRTLKGFQKDSFEKTLHVRREANIIMDDIGAEVTPW 242  
 DB 172 MAEELSERKGVSTLLHFPFAIDVKNALISSGYXDEIDAVKSVPIILDDIGAEQXTSW 231  
 QY 243 VRDEVGRLHYVNVHLPFPSSNPYSELEHNLATRGGEETKAARIIEVKSLSTP 302  
 DB 232 VRBEILQVLIQHRLLELPFTFSNYSFNDLERKMANIKGSDETQAKRVERVRYLAIE 291  
 QY 303 YFLSGENFR 311  
 DB 292 PHLEGPNRR 300

XX  
 OS Streptococcus pneumoniae type 4 strain.  
 XX  
 PN NO200277021-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 27-MAR-2002; 2002MO-IB02163.  
 XX  
 PR 27-MAR-2001; 2001GB-0007658.  
 XX  
 PA (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Maignani V, Tectelin H, Fraser C;  
 DR WPI: 2003-040579/03.  
 DR N-PSDB; ABX07476.  
 XX  
 PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
 PT useful as medicaments for treating or preventing a disease or infection  
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media  
 PT or ear infection -  
 PS  
 PS Claim 1; SEQ ID NO 3528; 56pp; English.  
 XX  
 CC The invention relates to a protein comprising or having at least 50%  
 CC identity to any of the 2469 amino acid sequences, identified in the  
 CC specification (available on a computer readable format), or its fragment,  
 CC expressed from 2469 of 2489 identified DNA coding regions from the  
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
 CC ABS5454. Also included are an antibody which binds one of the  
 CC proteins, treating a patient by administering the protein, DNA or  
 CC antibody (in a composition), a kit comprising first and second primers,  
 CC which are the nucleic acid cited above or fragments between nucleotides  
 CC 8-100 of a sequence not defined in the specification, for amplifying a  
 CC target sequence contained within a Streptococcus nucleic acid sequence,  
 CC where the first primer is substantially complementary to the target  
 CC sequence and the second primer is substantially complementary to the  
 CC complement of the target sequence, and where the parts of the primers  
 CC having substantial complementarity define the termini of the target  
 CC sequence to be amplified, assay comprising contacting a test compound  
 CC with the protein, and determining whether the test compound binds to the  
 CC protein and a Streptococcus pneumoniae bacterium, where one or more  
 CC genes encoding the proteins has been rendered inactive. The proteins,  
 CC nucleic acid molecules, antibody and compositions are useful as  
 CC medicaments for treating or preventing a disease or infection due to  
 CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,  
 CC sepsis, otitis media or ear infection. They are also useful in developing  
 CC vaccines, diagnostics and antibiotics. The methods are useful for  
 CC identifying immunodominant proteins. The present sequence is one of  
 CC the 2469 proteins expressed by the identified coding regions from the  
 CC genomic sequence.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 298 AA;  
 Query Match 23.9%; Score 392.5; DB 24; Length 298;  
 Best Local Similarity 32.3%; Pred. No. 5, 6e-27;  
 Matches 96; Conservative 64; Mismatches 116; Indels 21; Gaps 8;  
 QY 21 FEKRIEIKKEVINDPVKQPLEAHRALTNAMIDEDLVLOEY-KDQKHYDGHKFPDC 79  
 DB 17 YODLVQKIMK---DDVVAFTQ--QESLTPKELNSISKFNQYITERDKFLG---DT 66  
 QY 80 PNFVKGHVELVYNNRKIKIRYLCQPCCKIKYDEREALITSHMQRTLNAXKLDIYM 139  
 DB 67 DYIAKGYKYLKNNHGVADVSEETPELTAEEKAAIKRLKILNPASLKKASLAQVLD 126  
 QY 140 NHRDLVDV---AMAADICTAITNGEYQKGLYLGPFGTGKSFILGAIANQLSKK-VRS 195

Db 127 DDLGRLPVPEKLAFAVQYPAIR-----KGLYVGFQGVKSPMAVALAHNDSEKRGVS 181  
 QY 196 TTYLLEBFRTLLKGFQDSFEKKLRVREANILMDIDGAEVTPWTRDEVIGPLLR 255  
 Db 182 TLHAFSPFVLDVKNASDGNVKTLYVDEIKLSEVLIIDDIGABOSTWVRDEIIQVLIQVR 241  
 QY 256 MVHELPTFPSSNDYSELHHLMTRDGEKKT-KAARIRRVKSLSPYVLSGENFR 311  
 Db 242 MQENLPTFTSNFEDLEKHPAKVKGNDETWEARKVERIRYLABETRLBEGVNR 298

## RESULT 14

ABBS3757  
 ID ABB53757 standard; Protein; 291 AA.

AC ABB53757;

DT 16-MAY-2002 (first entry)

DE Lactococcus lactis protein pili4.

KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.

OS Lactococcus lactis IL1403.

PN FR2807446-A1.

PD 12-OCT-2001.

PF 11-APR-2000; 2000FR-0004630.

PR 11-APR-2000; 2000FR-0004630.

PA (INRG ) INRA INST NAT RECH AGRONOMIQUE.

PI Bolocline A, Sorokline A, Renault P, Ehrlich SD;

DR WPI; 2002-043418/06.

PT New nucleotide sequence useful in the identification or Lactococcus

PS Claim 6; SEQ ID No 459; 2504bp; French.

CC The present invention relates to a Lactococcus lactis nucleotide

CC sequence (ABA90521) and related proteins (ABB53300-ABB5621). The

CC nucleic acid sequence is useful in the detection and/or amplification of

CC nucleic acid sequence, particularly to identify Lactococcus lactis or

CC related species. The proteins of the invention are useful for the

CC biosynthesis or biodegradation of a composition of interest. The

CC invention helps research in lactic bacteria, particularly useful in the

CC production of yogurt and cheese.

CC Note: The sequence data for this patent is based on equivalent patent

CC WO20017734 (published 18-OCT-2001) which is available in electronic

CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 291 AA;

QY Query Match 6.8%; Score 112; DB 23; Length 291;

Db Best Local Similarity 21.8%; Pred. No. 0.088;

QY Matches 57; Conservative 36; Mismatches 94; Indels 74; Gaps 11;

Db 99 IRYLQCPCKIKYDE---ERFEALITIS-----HMQ-----RDTLN- 131

QY 29 VRX-KHPVSDYDPLIHGELVDGLQVTSMSYCTEACAKGIIYHQSSAETKANEDPLNK 87

QY 132 -----AKLNDIYVNHDRDLVYMAADIDICATNGSQVKGKYL 169

Db 88 SKYGKXSLTKQSLVGRKSLMPARFNTFKVNGEBOVNLQAORINAREYIQQGRFNTVFV 147

QY 170 YGPFGRKSFILGAIANQ-----LKSKEVRSITITLPEFIRTLKGFQDSFEKKLR- 222

Db 148 -CGAGRGKSHLMAILQEVNENIKDDKFTSLFINISELIREIKSNWNSDTAAEERLT 206  
 QY 223 -VRANILMLDIDIGAEV-----TPWRDDEVIGPLHVRVHLLPFFSSNPFYSEHNL 277  
 Db 207 LMRVLDLVIDLGESTFSPKSNWQ-----GVYINYNARGNITITSLTGKEM- 258  
 QY 278 AMTRDGEKTKAARIRRVKYS 298  
 Db 259 ---RSSYDKIVSRIMEGSKN 276

## RESULT 15

AAU38917  
 ID AAU38917 standard; Protein; 866 AA.

AC AAU38917;

DT 16-JAN-2002 (first entry)

DE C. pneumoniae CT113 homologue CPn0695.

KW Chlamydia; sexually transmitted disease; PID; antibacterial;

KW pelvic inflammatory disease; antigen; trachoma; gynecological;

KW acute respiratory tract infection; atherosclerosis; male infertility;

OS Chlamydia pneumoniae.

PN WO200181379-A2.

PD 01-NOV-2001.

PF 23-APR-2001; 2001WO-US13081.

PR 21-APR-2000; 2000US-198853P.

PR 20-JUL-2000; 2000US-219752P.

PA (CORI-) CORIXA CORP.

PI Bhatia A, Probst P, Stromberg EJ;

DR WPI; 2001-616771/71.

DR N-PSDB; AAS57049.

PT New polynucleotide for treating Chlamydia infections encodes a

PS polynucleotide containing an immunogenic portion of a Chlamydia

PS disclosure; Page 192-193; 208bp; English.

CC The invention relates to isolated polynucleotide encoding at least

CC a partial Chlamydia protein which is an antigenic fragment, or the

CC complements, fragments, homologues and variants, and antibodies

CC raised against the antigenic proteins (or fragments). The nucleic

CC acids, proteins and antibodies are used to diagnose and treat Chlamydia

CC infections (e.g. a sexually transmitted disease, pelvic inflammatory

CC disease (PID), acute respiratory tract infection, trachoma,

CC atherosclerosis and coronary heart disease) in a patient, and in

CC the treatment of male infertility. The compounds of the invention are

CC also useful for detecting the presence of Chlamydia in a patient, and

CC stimulating and/or expanding T cells specific for a Chlamydia protein.

XX The present sequence represents a Chlamydia antigen.

QY Query Match 6.7%; Score 110; DB 22; Length 866;

Db Best Local Similarity 22.0%; Pred. No. 0.56;

QY Matches 81; Conservative 71; Mismatches 123; Indels 94; Gaps 20;

Db 7 IMQFQSIINTSODEKRIEIKKEVINPDVQKPLEAHRAELTNMT--DEDLVNLOQY 64

QY 422 IVQGEAKIKRGQSPVQGEADAMQKSI-----DALREILASRLQGMDEKKLISGL 471

QY 65 KDQKHYDGHKADCPNFVKGHVPELYVDNNRI-KIRYLOCPC---KIKYDE----- 112  
Db 472 KEKNLSLESMKFSSE-----EBAERVADYVRVAELRYSLIPQLEBEIKQDEASLNQRDN 524  
QY 113 -----ERPEALLI---TSHHMQR--DTIAKLKDIYMHNRDL-----DVMAADDI 154  
Db 525 RLQOEYDERLIAQVANMTGI PVQKMLEGEAEKLLILEBSLEERVYVGOFPFAVSADSI 584  
QY 155 CTA---ITNGEVOVKGLIY--GPFGTGKSPFIGAIAIANOI KSKVSTIILPEF-----IR 205  
Db 585 RAARVGLANDPQRPGLGVFLFGPTGVGKTELAKALADLLFNKEAMVRFDMSEYMEKHSIS 644  
QY 206 TLKG-----GFKDGSPEKKLHVRVREANIIMLDDIGAEVTPWVRDEVIGPLHY----- 254  
Db 645 KLIGSSPGYVGYEBSLSLEALRRRYPYSVVLPDEIERKAD-----KEYLNTILLQVFPDDGI 698  
QY 255 -----RMVHELPTFF--SSNFDYSELHHLAMTRDGEKTKARITIERVKSLSPTY--- 303  
Db 699 LTDGKKRRVNCCKNALFTMTSNIGSPELADY--CSKKGSELTK-----EAILSVSPVLKR 751  
QY 304 FLGGENFRN 312  
Db 752 YLSPE-FMN 759

Search completed: December 15, 2003, 15:13:42  
Job time : 47.6369 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: December 15, 2003, 15:11:41 ; Search time 16.3604 Seconds  
(without alignments)  
809,474 Million cell updates/sec

Title: US-09-689-952-2  
Perfect score: 1643  
Sequence: 1 MGGGQSIMQFKSIINTSOD.....ERYKSLSTPYPLSGENFRNN 313

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/aa/5A COMB pep:\*  
2: /cgn2\_6/ptodata/1/aa/5B COMB pep:\*  
3: /cgn2\_6/ptodata/1/aa/6A COMB pep:\*  
4: /cgn2\_6/ptodata/1/aa/6B COMB pep:\*  
5: /cgn2\_6/ptodata/1/aa/PTUS COMB pep:\*  
6: /cgn2\_6/ptodata/1/aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1643	100.0	313	4 US-09-470-512A-2	Sequence 2, Appl1
2	1630	99.2	313	4 US-09-470-512A-10	Sequence 10, Appl1
3	1378	83.9	307	4 US-09-134-001C-3792	Sequence 3792, Ap
4	586	35.7	280	4 US-09-470-512A-14	Sequence 14, Appl1
5	568	34.6	278	4 US-09-470-512A-15	Sequence 15, Appl1
6	449	27.3	310	4 US-09-107-532A-5560	Sequence 5560, Ap
7	124	7.5	268	4 US-09-328-352-5963	Sequence 5963, Ap
8	110	6.7	872	4 US-09-198-452A-163	Sequence 163, App
9	105	6.4	1866	4 US-08-938-105-3	Sequence 3, Appl1
10	102.5	6.2	304	4 US-09-134-001C-5153	Sequence 5153, Ap
11	99.5	6.1	443	4 US-09-134-001C-4211	Sequence 4211, Ap
12	98.5	6.0	467	4 US-09-198-452A-443	Sequence 443, App
13	98	6.0	1964	2 US-08-790-912-3	Sequence 3, Appl1
14	98	6.0	2052	2 US-08-790-912-2	Sequence 2, Appl1
15	97	5.9	389	2 US-08-820-170A-13	Sequence 13, Appl1
16	97	5.9	388	3 US-09-055-699-13	Sequence 13, Appl1
17	97	5.9	389	3 US-09-273-565-13	Sequence 13, Appl1
18	97	5.9	389	3 US-09-565-538-13	Sequence 13, Appl1
19	97	5.9	389	4 US-09-661-468-13	Sequence 13, Appl1
20	97	5.9	389	4 US-09-976-165-13	Sequence 13, Appl1
21	96.5	5.9	406	2 US-08-232-719-1	Sequence 1, Appl1
22	96.5	5.9	406	2 US-08-470-925-1	Sequence 1, Appl1
23	96.5	5.9	406	2 US-08-471-613-1	Sequence 1, Appl1
24	96.5	5.9	406	5 PCT-US93-10443-1	Sequence 1, Appl1
25	95.5	5.8	956	4 US-08-559-896B-2	Sequence 2, Appl1
26	95.5	5.8	956	4 US-09-134-001C-4452	Sequence 4452, Ap
27	94.5	5.8	942	4 US-09-254-776B-51	Sequence 51, Appl1

28	94	5.7	793	2 US-08-468-558-5	Sequence 5, Appl1
29	94	5.7	793	3 US-08-676-444-5	Sequence 5, Appl1
30	92	5.6	312	4 US-09-134-001C-3676	Sequence 3676, Ap
31	92	5.6	468	4 US-09-134-001C-5156	Sequence 5156, Ap
32	92	5.6	562	4 US-09-328-352-7898	Sequence 7898, Ap
33	92	5.6	1939	4 US-09-310-187A-1	Sequence 1, Appl1
34	91.5	5.6	803	4 US-09-154-750A-85	Sequence 85, Appl1
35	91	5.5	761	4 US-09-328-352-7492	Sequence 7492, Ap
36	90.5	5.5	859	4 US-09-199-637A-281	Sequence 281, App
37	90.5	5.5	859	4 US-09-252-991A-21413	Sequence 21413, A
38	90.5	5.5	878	4 US-09-134-001C-4378	Sequence 4378, Ap
39	90	5.5	405	2 US-08-222-719-2	Sequence 2, Appl1
40	90	5.5	405	2 US-08-470-925-2	Sequence 2, Appl1
41	90	5.5	405	2 US-08-471-613-2	Sequence 2, Appl1
42	90	5.5	405	5 PCT-US93-10443-2	Sequence 2, Appl1
43	90	5.5	647	4 US-09-134-001C-5458	Sequence 5458, Ap
44	89.5	5.4	717	4 US-08-924-629C-5	Sequence 5, Appl1
45	89.5	5.4	863	4 US-09-328-352-6730	Sequence 6730, Ap

## ALIGNMENTS

RESULT 1  
US-09-470-512A-2  
; Sequence 2, Application US/09470512A  
; Patent No. 6376652  
; GENERAL INFORMATION:  
; APPLICANT: PhageTech, Inc.  
; TITLE OF INVENTION: Compositions and methods involving an essential Staphylococcus Au  
; FILE REFERENCE: 21715/1000  
; CURRENT APPLICATION NUMBER: US/09/470,512A  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 313  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-470-512A-2

Query Match 100.0%; Score 1643; DB 4; Length 313;  
Best Local Similarity 100.0%; Pred. No. 1.3e-155;  
Matches 313; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGGGQSIMQFKSIINTSODFEKRIKKEVINDDVQFLEAHRALTNMIDEDLV 60  
DB 1 MGGGQSIMQFKSIINTSODFEKRIKKEVINDDVQFLEAHRALTNMIDEDLV 60

QY 61 LQRYKQOQKHVDKRAQDCPNFYKGVPELYVNNRIKIRYLOCPCKIKYDERFEAEI 120  
DB 61 LQRYKQOQKHVDKRAQDCPNFYKGVPELYVNNRIKIRYLOCPCKIKYDERFEAEI 120

QY 121 TSHHMORDTLNAXLKDQIYNNHRDLDPVMAADICTAITNGEQQVGLYLGPGTGKSF 180  
DB 121 TSHHMORDTLNAXLKDQIYNNHRDLDPVMAADICTAITNGEQQVGLYLGPGTGKSF 180

QY 181 LGAIANQLSKKVRSITIIYLPFRITLKGFGKGSFEKLIHVRVANIIMLDIGAEV 240  
DB 181 LGAIANQLSKKVRSITIIYLPFRITLKGFGKGSFEKLIHVRVANIIMLDIGAEV 240

QY 241 PWRDEVIGPLHRYRWHLPTEFSNPNYSELEHLLAMTRDDEEKTARITERYVSL 300  
DB 241 PWRDEVIGPLHRYRWHLPTEFSNPNYSELEHLLAMTRDDEEKTARITERYVSL 300

QY 301 TPYFLSGENFRNN 313  
DB 301 TPYFLSGENFRNN 313

RESULT 2  
US-09-470-512A-10

```

: Sequence 10: Application US/09470512A
: Patent No. 6376652
: GENERAL INFORMATION:
: APPLICANT: Phageotech, Inc.
: TITLE OF INVENTION: Compositions and methods involving an essential Staphylococcus Au
: TITLE OF INVENTION: gene and its encoded protein
: FILE REFERENCE: 21715/1000
: CURRENT APPLICATION NUMBER: US/09/470,512A
: CURRENT FILING DATE: 1999-12-12
: NUMBER OF SEQ. ID NOS: 16
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 10
: LENGTH: 313
: TYPE: PRN
: ORGANISM: Staphylococcus aureus
US-09-470-512A-10

```

Query Match	99.2%	Score 1630;	DB 4;	Length 313;
Best Local Similarity	98.7%;	Pred. No. 2.5e-154;		
Matches 309;	Conservative 4;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	1	MGGSJIMQFKSIINTSODFEKRIEKKIYINPDVKQFLEAHRELTLAMIDELNV	60
Db	1	MGGSJIMQFKSIINTSODFEKRIEKKIYINPDVKQFLEAHRELTLAMIDELNV	60
Qy	61	LOEYKQOQXHYDHKRAADCPNVKGVHPELYDNNRIKIRLYOCPCKIKUDEEFAELI	120
Db	61	LOEYKQOQXHYDHKRAADCPNVKGVHPELYDNNRIKIRLYOCPCKIKUDEEFAELI	120
Qy	121	TSHHMORDTLNAKLKQIYVNNHRDRLDVMAAADDICTAITNGEQVKGYLVPFSTGKSF	180
Db	121	TSHHMORDTLNAKLKQIYVNNHRDRLDVMAAADDICTAITNGEQVKGYLVPFSTGKSF	180
Qy	181	LGAIANQLKSKKVRSTIYLPFPITLKGFPDGSFEKKLHRVRBANIMLMDIGAEVYT	240
Db	181	LGAIANQLKSKKVRSTIYLPFPITLKGFPDGSFEKKLHRVRBANIMLMDIGAEVYT	240
Qy	241	PWVRDEVIGPLLYHRVWHELPTFFSSNPDYSLEHNLMTADGEEKTKAARIIRVSVLS	300
Db	241	PWVRDEVIGPLLYHRVWHELPTFFSSNPDYSLEHNLMTADGEEKTKAARIIRVSVLS	300
Qy	301	TPYFLSGENFRNN	313
Db	301	TPYFLSGENFRNN	313

```

RESULT 3
US-09-134-001C-3792
; Sequence 3792, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lym Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3792
; LENGTH: 307
; TYPE: PR1
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-3792

```

Query Match	83.9%	Score 1378;	DB 4;	Length 307;
Best Local Similarity	82.7%;	Pred. No. 3.2e-129;		
Matches 254;	Conservative 34;	Mismatches 19;	Indels 0;	Gaps 0;
7 IMKQFSINTSQDFEKRIEIKKEVINDPVQGTLEHRAELTNAMIDEDDLNTVQEKD 66				

[illegible]

```

RESULT 4
US-09-470-512A-14
; Sequence 14, Application US/09470512A
; Patent No. 6376652
; GENERAL INFORMATION:
; APPLICANT: PhageTech, Inc.
; TITLE OF INVENTION: Compositions and methods involving an essential Staphylococcus Aureus protein
; TITLE OF INVENTION: gene and its encoded protein
; FILE REFERENCE: 21715/1000
; CURRENT APPLICATION NUMBER: US/09/470,512A
; CURRENT FILING DATE: 1999-12-12
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 14
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-470-512A-14

```

	Query Match	35.74; Score 586; DB 4; Length 280;
	Best Local Similarity 42.9%; Pred. No. 2.4e-50;	
	Matches 120; Conservative 57; Mismatches 101; Indels 2; Gaps 2;	
QY	35 DEDYKQPLFAHRAALITLAAIMDEDLNAVLTQYKDQOKYHD-GHKFADCCPFVGVGHAPELYVD	93
DB	1 DQDVAQPLKENEVEVIDDKMT EKSLINKIYEYEBQSKNCSYSEDNCNNLLTGYPHKLAVN	60
QY	94 NNRIRKIRLOCPCKIKYDEERFEAEALITSHHQBDTLNAKIKDIYMHNRDRLDVAAMAAD	153
DB	61 GRSDIDIEYBECPPVVRKDKDQCKQKQSLKMSWYIQGDDLGATGQYVIDSPSRLLAMFQHYTD	120
QY	154 ICTAITNGEQVKGILYLYGPGTGSFLIGAIANOLKSKKARSTIILYLPFRITLKGCKYD	213
DB	121 FLKSYNETGKKKGILYLKGFQVGTGFMFLAALANLAEKEYSMTLYVYDEPRELKNLSLQD	180
QY	214 GSPEFKILRYVEANILMLDIDGAEVTPWVDEVIYGLPLIHRVYHELPFFSSNFDYSEL	273
DB	181 QTLBEKILNMVKTITVLMMLDDIGASMSWVRDEVIYGLVQLQHRMGQQLPTFFSSNPSPEL	240
QY	274 EHHILAMTRDGE-EKTKAARITIERKYSISTPFLSGENFRN	312
DB	241 KGHFTVYSORGEKEVKAARIMERTILYLAAPRLIDGENRRH	280

```

RESULT 5
US-09-470-512A-15
; Sequence 15, Application US/09470512A
; Patent No. 6376652
; GENERAL INFORMATION:

```

APPLICANT: PhageTech, Inc.  
 TITLE OF INVENTION: Compositions and methods involving an essential Staphylococcus Au  
 TITLE OF INVENTION: gene and its encoded protein  
 FILE REFERENCE: 21715/1000  
 CURRENT APPLICATION NUMBER: US/09/470,512A  
 CURRENT FILING DATE: 1999-12-12  
 NUMBER OF SEQ ID NOS: 16  
 SOFTWARE: PatentIn version 3.0  
 SEQ ID NO 15  
 LENGTH: 278  
 TYPE: PRF  
 ORGANISM: Bacillus halodurans  
 US-09-470-512A-15

Query Match 34.6%; Score 568; DB 4; Length 278;  
 Best Local Similarity 40.6%; Pred. No. 1.4e-48;  
 Matches 117; Conservative 61; Mismatches 90; Indels 20; Gaps 7;

QY 36 PVKQPLEAHRAALTNAMIDEDLVNLYQYKDOQKHVDG--KPADQPNFVKGHVPELYAD 93  
 DB 1 PNYQFLBEH-PSLSITTEBGLSKLYEQKQSHC-AHCPGLQCKPMLKGTQPTLYE 58  
 QY 94 NNRKIRYLQCPCKIKYDERFEAEITSHMORDTLNAKLDI-----YNNHRDRLD 146  
 DB 59 RDLSELSTPCPLKEEEREKKKSLIRSLYIPKEILEAKFDVESEPGRSIASHR-ALE 117  
 QY 147 VMAADICTALTNGQVGLYVYGPFGKSFILGAINOLKSKKVRSTIYLPFFIT 206  
 DB 118 FALSAP-----GDDGGLYLYGKFGVGTFLMGAINELMDRDISITIVVDPFPE 170  
 QY 207 LKGFQDGFSEKGLHYVREANILMLDIGAEEVTPMVRDEVIGPLHYHVMHLPFFSS 266  
 DB 171 LKQSIDGTFQQLDVKQAQVILFDIGLETWTSVRDVLVLIQYHMEGLPTLFTS 230  
 QY 267 NPDYSELHHLAMT--RDGEKTKAARIIEKVKSLSTPYFLSGENFNN 313  
 DB 231 NVDYDELEHIAVNDKSGTELLAKKVRMERIRHYTYSVMVQGNREH 278

RESULT 6  
 US-09-107-532A-5580  
 Sequence 5580, Application US/09107532A  
 Patent No. 6583275  
 GENERAL INFORMATION:  
 APPLICANT: Lynn A Doucette-Stamm and David Bush  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
 ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
 NUMBER OF SEQUENCES: 7310  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
 STREET: 100 Beaver Street  
 CITY: Waltham  
 STATE: Massachusetts  
 COUNTRY: USA  
 ZIP: 02354  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: CD/ROM ISO9660  
 COMPUTER: PC  
 OPERATING SYSTEM: <Unknown>  
 SOFTWARE: ASCII  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/107,532A  
 FILING DATE: 30-Jun-1998  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 60/085,598  
 FILING DATE: 14 May 1998  
 APPLICATION NUMBER: 60/051571  
 FILING DATE: July 2, 1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Arinello, Pamela Deneka  
 REGISTRATION NUMBER: 40,489  
 REFERENCE/DOCKET NUMBER: GTC-012  
 TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007  
 TELEFAX: (781)893-8277  
 INFORMATION FOR SEQ ID NO: 5580:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 310 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: YES  
 ORIGINAL SOURCE:  
 ORGANISM: Enterococcus faecium  
 FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: (8) LOCATION 1...310  
 SEQUENCE DESCRIPTION: SEQ ID NO: 5580:  
 US-09-107-532A-5580

Query Match 27.3%; Score 449; DB 4; Length 310;  
 Best Local Similarity 33.5%; Pred. No. 1.2e-36;  
 Matches 106; Conservative 69; Mismatches 123; Indels 18; Gaps 7;

QY 5 GSIMQFSTINTSQDFEKRIRKIKYVNDPVKQFLFAHRAELTNAMIDEDLVNLYQY 64  
 DB 4 EDVGKELKRII--SKNYQEKYAEMLAEVINDADVAFLSAHQEBILTQADIERKSYAKLYEF 62  
 QY 65 KQQKHGDSKRPADCPNPFVKGHVPELYDNNRIRKIRYLQCPCKIKYDERFEAEILTSHH 124  
 DB 63 VQEKRYQKRN--DPTMAIPGTEPKLTTFHVDVYVTEMLARQOEIRSRKAKND 119  
 QY 125 MORDTLNAKLDIYNNHRDRLD-----VMAADIC--TAINSGQVGLYLYGPGTK 177  
 DB 120 MKDQISAFENF-----ERTDGRGASLEALDPEKTYADPKGYH-KGLYLYGSGICK 173  
 QY 178 SFILGAINOLKSKKVRSTIYLPFFITRLKGFQDGFSEKGLHYVREANILMLDIGAE 237  
 DB 174 TYLGATRAELAIAGFTTLVHFPTFAVEMKQAIQDQVAEKLDVAKRSPILMIDIGAD 233  
 QY 238 EYTPMVRDEVIGPLHYHVMHLPFFSSNPFYSELEHHLAMTRDQ--EKKTAARIIEKV 296  
 DB 234 AMSSWIRDEVFVLIQYRQEBOLPTFFSSNFTMDELOHLSVYQRDREBPLAKRIMERI 293  
 QY 297 KSLSTPYFLSGENFNN 312  
 DB 294 KYLTKEIENTGRNRN 309

RESULT 7  
 US-09-328-352-5963  
 Sequence 5963, Application US/09328352  
 Patent No. 6562958  
 GENERAL INFORMATION:  
 APPLICANT: Gary L. Breton et al.  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
 FILE REFERENCE: GTC99-03PA  
 CURRENT APPLICATION NUMBER: US/09/328,352  
 CURRENT FILING DATE: 1999-06-04  
 NUMBER OF SEQ ID NOS: 8252  
 SEQ ID NO 5963  
 LENGTH: 268  
 TYPE: PRF  
 ORGANISM: Acinetobacter baumannii  
 US-09-328-352-5963

Query Match 7.5%; Score 124; DB 4; Length 268;  
 Best Local Similarity 25.6%; Pred. No. 0.00024;  
 Matches 34; Conservative 28; Mismatches 63; Indels 8; Gaps 4;

QY 107 KIKYDERFEAEILTSHMORDTLNAKLDIYNNHRDRLDVMAADICTAIT--NGRQ 163  
 DB 102 KYRBEPEVANTVAVSSKSKQKLTNPQTFSLFVEGRSN--QMAAETGRKVLTOIGASQ 158  
 QY 164 VAGLYLYGPGFGTKSFILGAINOLKSKKVRSTIYLP--PEFLRTLKGFQDGFSEKGLH 221

Db 159 HNPFLYGPFGTGLKTHLMQAVGNALLQAKPNARVMTVSQSPVQDFVSLQKGVVEFFK 218  
Qy 222 RVEANIMLMDI 234  
Db 219 NCRSLDLVLVDI 231

## RESULT 8

US-09-198-452A-163  
Sequence 163, Application US/09198452A  
Patent No. 6559294  
GENERAL INFORMATION:  
APPLICANT: Grifffais, R.  
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection  
TITLE OF INVENTION: and treatment of infection  
FILE REFERENCE: 9710-003-999  
CURRENT APPLICATION NUMBER: US/09/198,452A  
CURRENT FILING DATE: 1998-11-24  
NUMBER OF SEQ ID NOS: 6849  
SEQ ID NO 163  
LENGTH: 872  
TYPE: PRT  
ORGANISM: Chlamydia pneumoniae  
US-09-198-452A-163

Query Match 6.7%; Score 110; DB 4; Length 872;  
Best Local Similarity 22.0%; Pred. No. 0.034;  
Matches 81; Conservative 71; Mismatches 123; Indels 94; Gaps 20;

Qy 7 IMQFQKINTSGDFEKKRIKKEVINDPVQFLAHRALTNAMI--DEDLNVLQVY 64  
Db 428 IVKQEARIKRQSPSYQREADLMQKSI-----DALREELASLRLGMDKEKGLISGL 477  
Qy 65 KDQKHYDHGKADCPFVKGHVDELVDNRI-KIRYLOQPC--KIKYDE----- 112  
Db 478 KKKKNSLESKFS-----EEMERVADYRVAELVSLIPQLEBEIKQDESLNQRDN 530  
Qy 113 -----ERPEALIT--TSHMQR--DTLNAKLKDIYNNHRDL-----DVAMADDI 154  
Db 531 RLLOEYVERLQAVVNMGTGIPVQKMLEGEAEKLLLESLEBERVGGPFAVSADSI 590  
Qy 155 CTA---TTNGEYVGLVY-GPFGTGSFLIGAIANOLKSKKVVSTIIVPEF-----IR 205  
Db 591 RAARVGLNDPQRLGVFLFGPTGVGKTELAKALDLLENKEMARPDSEVMEKKSIS 650  
Qy 206 TLKG-----GFKDGFPEKRLHVRVREANIMLDDIGAEVTPVVRDEVIGPLHY----- 254  
Db 651 KLIGSSPGVGYEBSGSLSEALRRPYSVVLFDIEIKAD-----KEVLNILLQVFDGI 704  
Qy 255 -----RMVHELPTFP--SSNFYSELHHLAMTRDGEKTKARITERYKSLSTPY--- 303  
Db 705 LTDDKKRKVCKNALFTMTSNIGSPELADY--CSKKSSELTK-----EALISVSPVLKR 757  
Qy 304 FLGSEFRN 312  
Db 758 YLSPE-FRN 765

## RESULT 9

US-08-938-105-3  
Sequence 3, Application US/08938105  
Patent No. 6353151  
GENERAL INFORMATION:  
APPLICANT: Leinwand, Leslie A.  
TITLE OF INVENTION: Viketrom, Karen L.  
TITLE OF INVENTION: TRANSGENIC MODEL FOR HEART FAILURE  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross P.C.  
STREET: 1700 Lincoln St., Suite 3500  
CITY: Denver

STATE: CO  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/938,105  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Crook, Marnell M.  
REGISTRATION NUMBER: 31,071  
REFERENCE/DOCKET NUMBER: 3595-4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1886 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULAR TYPE: protein  
US-08-938-105-3

Query Match 6.4%; Score 105; DB 4; Length 1886;  
Best Local Similarity 24.9%; Pred. No. 0.33;  
Matches 58; Conservative 24; Mismatches 79; Indels 72; Gaps 10;

Qy 18 SDFEKKRIKKEVINDPVQFLAHRALTNAMIDEDLVLYQVQKQKHVDGK-- 75  
Db 791 SAATEKEMAMKEFGVYDALEKSPARKLEKRV-----SLQKNDQLQVQAEQDN 846  
Qy 76 PAD-----CPNFVKGHVDELVDNRIKIRYLOQPCRK-----YDERPEALITSHM 125  
Db 847 LADAERCDQLK-----NKI-----QLEAKVKEWTERLEDEEMNAALTLAKRK 891  
Qy 126 QRTLNAKLKDIYNNHRDLVDVMAADDICTALTNGEYVGLVYGPFGTGSFLIGATA 165  
Db 892 LBECESELDKI-----DDELTTLAKVKEKH-----A 919  
Qy 186 NOLSKKVVSTIIVPEFIRTLKGFGDGSFEKKLHVRVREANIMLDDIGAE 238  
Db 920 TENKVNLTTEMAGLDEITAKL-----TKKK--ALQEAHQALDDLAEE 963

## RESULT 10

US-09-134-001C-5153  
Sequence 5153, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 5153  
LENGTH: 304  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5153

Query Match 6.2%; Score 102.5; DB 4; Length 304;  
Best Local Similarity 21.4%; Pred. No. 0.04;  
Matches 33; Conservative 35; Mismatches 59; Indels 27; Gaps 4;

QY 151 ADDICTATNGEVOVKGLYVPGTGKSPFLGAIANOLSKKYRSTIYLPE-----EFL 204  
 DB 135 ASLAVAEAPAEVYNPLFTIGVGLGKTHLMHAIGHVSNKNAYITSSSEKFTNERT 194  
 QY 205 RTLKGGFKDGSPEKCLHRYVEANILMDDIGAEVTPWVDEVIPLHYRMVHELPTFF 264  
 DB 195 KSI-----RDNETAFAFEKTKRIDVLLIDDIQFIONNEQOQE-----PF 234  
 QY 265 SSNFYSELEHMLAMTRDGEKTKAARIIEKVS 298  
 DB 235 HFMELHONNKQIVISSDRPPK-EIAKLEDRIRS 267

## RESULT 11

US-09-134-001C-4211  
 ; Sequence 4211, Application US/09134001C  
 ; Patent No. 6380370  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 ; FILE REFERENCE: GTC-007  
 ; CURRENT APPLICATION NUMBER: US/09/134,001C  
 ; PRIOR FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064,964  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5674  
 ; SEQ ID NO 4211  
 ; LENGTH: 443  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus epidermidis  
 ; US-09-134-001C-4211

Query Match 6.1%; Score 99.5; DB 4; Length 443;  
 Best Local Similarity 20.4%; Pred. No. 0.14;  
 Matches 57; Conservative 58; Mismatches 115; Indels 49; Gaps 13;

QY 1 MGGGGSIMKQFKSI-----INTSQDEKRIEIKKEVINPDVKQFLAHRALTNAMTD 55  
 DB 6 MPCRERTSPQYESFHELYKNNTTKELQKXSKLNTYSKLNKBLVLA-----IMEAQWE 61  
 QY 56 EDLANTQETKQDQKATDGHKACDCEPVFKCHVELYVNNRRIKIRYLQCCCKI-----K 109  
 DB 62 KQGNVYMEGLDIDODGVEFLRTVYVSKGE-KDIYISASQIRFBIKRGDKVTKGVKRP 120  
 QY 110 YDEERF-----EALITSHMQ-----RDTLNAKLKDIYNNRDLV-----AQAADICT 156  
 DB 121 KONEKTYGLLOVDFVNDHNAEYKAPHPQA-LTPLYPERIILETOSTYSTRIMDLVT 179  
 QY 157 AITNGEVOVKGLYVPGTGKSPFLGAIANOLSKKYRSTIYL-----PEFIRTLKGGF 211  
 DB 180 PIGLGG--RGL-IVAPPKAGKTSLLKEINAIASNNPEANLFLVLGGEREEVTDIERSV 236  
 QY 212 KQGSF-----EKKLHRYVEANILM-----LDDIGAEV 239  
 DB 237 BSAEVAHSTFDEBPBHHVKAVALILBRAKLVEIGEDVI 275

## RESULT 12

US-09-198-452A-443  
 ; Sequence 443, Application US/09198452A  
 ; Patent No. 6559294  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Griffiths, R.  
 ; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments  
 ; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention  
 ; FILE REFERENCE: 9710-003-999  
 ; CURRENT APPLICATION NUMBER: US/09/198,452A  
 ; CURRENT FILING DATE: 1998-11-24

NUMBER OF SEQ ID NOS: 6849  
 ; SEQ ID NO 443  
 ; LENGTH: 467  
 ; TYPE: PRT  
 ; ORGANISM: Chlamydia pneumoniae  
 ; US-09-198-452A-443

Query Match 6.0%; Score 98.5; DB 4; Length 467;  
 Best Local Similarity 19.1%; Pred. No. 0.19;  
 Matches 43; Conservative 49; Mismatches 90; Indels 43; Gaps 9;

QY 39 KQFLAHRALTNAMIDEDLVLYQBYKQDQKQYDGHKACDCEPVFKCHVELYVNNRIRK 98  
 DB 28 EQFLNVTKRGSKTAPEMWSIPQVLEEQEK-----IRLEVPNIFVQVLYLD 75  
 QY 99 IRYLOPCK-IKYDERFERA-ELITSRH-----MORDTLNAKLKDIYNNRDL-----RL 145  
 DB 76 -NYKRDLSGFVPLDYGEPALFEVVAEHKKPAPVAVSOKES-NEGISEFEEETKQFELTL 133  
 QY 146 DYAMADICTATN-----GEQVKGVLVPGTGKSPFLGAIANOLKS- 190  
 DB 134 NUSYRDNFTBSPNQFVKSAAVGIAGKGRSYNPLFTIGVGLGKTHLMHAIGHVYRER 193  
 QY 191 -KKVSTIYLPEFIRTLKGGFKDGSPEKCLHRYVEANILMDDI 234  
 DB 194 HKNLRHCTTEAFINDLVYHLKSKSVDMKNRFPYSLDLVDI 238

## RESULT 13

US-08-790-912-3  
 ; Sequence 3, Application US/08790912  
 ; Patent No. 5976542

GENERAL INFORMATION:  
 APPLICANT: Weiser, Jeffrey N.  
 APPLICANT: Plaut, Andrew G.  
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT  
 TITLE OF INVENTION: OF STREPTOCOCCUS PNEUMONIAE INFECTION  
 NUMBER OF SEQUENCES: 10  
 CORRESPONDENCE ADDRESS:  
 ADDRESSES: PANITCH SCHWARZE JACOBS & NADEL, P.C.  
 STREET: 1601 Market Street, 36th Floor  
 CITY: Philadelphia  
 STATE: Pennsylvania  
 COUNTRY: USA  
 ZIP: 19103-2398  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/790,912  
 FILING DATE: 29-JAN-1997  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 60/026,939  
 FILING DATE: 23-SEP-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Leary, Kathryn  
 REGISTRATION NUMBER: 36,317  
 REFERENCE/DOCKET NUMBER: 7600-401  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (215) 567-2020  
 TELEFAX: (215) 567-2991  
 TELEX: 831-494  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1964 amino acids  
 TYPE: amino acid  
 STRANDBESS:  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-790-912-3

Query Match 6.0%; Score 98; DB 2; Length 1964;  
 Best Local Similarity 20.7%; Pred. No. 1.8;  
 Matches 75; Conservative 53; Mismatches 110; Indels 124; Gaps 20;

QY 23 KRIEKT-----KEVINDDP-VKQFLHRAELTNMIDEDLVLOEYDQOK-----69  
 DB 1124 KNIKELLPYNNKSTIVKYGVLVENSLLYQKELSAVMMKDDQVITDIYNNKOTANKLL 1183  
 QY 70 HYDGHK-----FADCPNFKGHPVLYDNNRIKIRYLQCPCKIKYBERFEAE 118  
 DB 1184 HYNDHSEKFDLKYQDFPANNPEYNNGLTYTPN-----QFLYRDSIVKE 1231  
 QY 119 LITSHHMQ-----RDTL-----NAKLDIYM-----NHRDL-----DVAMA 150  
 DB 1232 VLP--ELQKLDYQSDAIRKTLGISPEVKLTLEYLEDQFSKTKONLGSLSKLSADAGLA 1289  
 QY 151 ADDICTAINTGBOVKGLYLGPGTGKSPILGAIAN-----QLSKKYRSTIITYLPER 203  
 DB 1290 SDN---SVTRG-----YLVDKIKNNKEALLGLTYLERWYNNFYQVNVKDLVMYHPDF 1340  
 QY 204 -----IRTLKGFKDGSFEKKLHVRVANILM---LDDIGAEVTPWVRDEVI 248  
 DB 1341 FGKGNTPDLTLBELKSGFN-----NLLAKNNVDYTGISLSAQHGATDLF 1386  
 QY 249 GPLHRYRMVHELPTFFSSNPDYSELEHHLAMTRDGEKTKAARIIRVXSLSTPYPLSGE 308  
 DB 1387 STLEHRYKVF-LPNTSNNDWFKSETKAYIV-----EKS---TIEBVK---TKQGLAGT 1433  
 QY 309 NF 310  
 DB 1434 KY 1435

RESULT 14  
 US-08-790-912-2  
 Sequence 2, Application US/08790912  
 Patent No. 5976542

GENERAL INFORMATION:  
 APPLICANT: Weiser, Jeffrey N.  
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT  
 TITLE OF INVENTION: OF STREPTOCOCCUS PNEUMONIAE INFECTION  
 NUMBER OF SEQUENCES: 10  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: PAINTCH SCHWARZE JACOBS & NADEL, P.C.  
 STREET: 1601 Market Street, 36th Floor  
 CITY: Philadelphia  
 STATE: Pennsylvania  
 COUNTRY: USA  
 ZIP: 19103-2398  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/790,912  
 FILING DATE: 29-JAN-1997  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 60/026,939  
 FILING DATE: 23-SEP-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Leary, Kathryn  
 REGISTRATION NUMBER: 36,317  
 REFERENCE/DOCKET NUMBER: 7600-401  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (215) 567-2020  
 TELEFAX: (215) 567-2991  
 TELEX: 831-494  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:

LENGTH: 2052 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-790-912-2

Query Match 6.0%; Score 98; DB 2; Length 2052;  
 Best Local Similarity 20.7%; Pred. No. 1.9;  
 Matches 75; Conservative 53; Mismatches 110; Indels 124; Gaps 20;

QY 23 KRIEKT-----KEVINDDP-VKQFLHRAELTNMIDEDLVLOEYDQOK-----69  
 DB 1197 KNIKELLPYNNKSTIVKYGVLVENSLLYQKELSAVMMKDDQVITDIYNNKOTANKLL 1256  
 QY 70 HYDGHK-----FADCPNFKGHPVLYDNNRIKIRYLQCPCKIKYBERFEAE 118  
 DB 1257 HYNDHSEKFDLKYQDFPANNPEYNNGLTYTPN-----QFLYRDSIVKE 1304  
 QY 119 LITSHHMQ-----RDTL-----NAKLDIYM-----NHRDL-----DVAMA 150  
 DB 1305 VLP--ELQKLDYQSDAIRKTLGISPEVKLTLEYLEDQFSKTKONLGSLSKLSADAGLA 1362  
 QY 151 ADDICTAINTGBOVKGLYLGPGTGKSPILGAIAN-----QLSKKYRSTIITYLPER 203  
 DB 1363 SDN---SVTRG-----YLVDKIKNNKEALLGLTYLERWYNNFYQVNVKDLVMYHPDF 1413  
 QY 204 -----IRTLKGFKDGSFEKKLHVRVANILM---LDDIGAEVTPWVRDEVI 248  
 DB 1414 FGKGNTPDLTLBELKSGFN-----NLLAKNNVDYTGISLSAQHGATDLF 1459  
 QY 249 GPLHRYRMVHELPTFFSSNPDYSELEHHLAMTRDGEKTKAARIIRVXSLSTPYPLSGE 308  
 DB 1460 STLEHRYKVF-LPNTSNNDWFKSETKAYIV-----EKS---TIEBVK---TKQGLAGT 1506  
 QY 309 NF 310  
 DB 1507 KY 1508

RESULT 15  
 US-08-820-170A-13  
 Sequence 13, Application US/08820170A  
 Patent No. 5831058

GENERAL INFORMATION:  
 APPLICANT: Teutomu, FUJIWARA  
 APPLICANT: Takeshi, WATANABE  
 APPLICANT: Masato, HORIE  
 TITLE OF INVENTION: HUMAN GENE  
 NUMBER OF SEQUENCES: 42  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Sughrue, Mion, Zimm, Macpeak & Seas  
 STREET: 2100 Pennsylvania Avenue, N.W.  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: United States  
 ZIP: 20037-3202  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/820,170A  
 FILING DATE:  
 CLASSIFICATION: 536  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 293-7060  
 TELEFAX: (202) 293-7860  
 TELEX: 6491103  
 INFORMATION FOR SEQ ID NO: 13:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 389 amino acids

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-820-170A-13

Query Match 5.9%; Score 97; DB 2; Length 389;  
Best Local Similarity 23.9%; Pred. No. 0.2;  
Matches 62; Conservative 39; Mismatches 86; Indels 72; Gaps 14;

QY	19 QPFEKRIEKKVAVNDPVKQFLFAHRAELTN--AMIDELANVLOE---YKDOQKHVD 72
DB	10 QDYRKQLLE-HKEI--DGRLEKELEQIK-ELTKQYKESNDLKALQSVGQIVGEVLKQLT 65
QY	73 GHKF---ADCPVFVKGHPVELVVDNNRIKIRYLQCPCKIKYDEERFEALITSHMQRD 128
DB	66 EEKPIVKAATNGPRVVG-----CRQLDKSKLKPGTRVALDPTTL 105
QY	129 TLNAKLKD-----LY-MNHRDLRLDVAMA-----ADDICTAINTNG----- 162
DB	106 TLMRYLPRREVDPLVYMNMSHEDPGNVSYSEIGLSEQIRELREVIELPLTNPELFORVGII 165
QY	163 QVKGILYGPFGTGKSFILGAINOLKS---KKVSTII--YLPFIRTLKGGFKDGSFE 217
DB	166 PPKGCLLYGPPGTGKTLAAVAASQDCNFKVSSIVDKYIGESARLIREMF----- 219
QY	218 KKLHVRVREANIIMLDDIGA 236
DB	220 -NYARDHPCCILFMDSIDA 237

Search completed: December 15, 2003, 15:17:39  
Job time : 18.3604 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 15:13:51 / Search time 30.5681 Seconds  
(without alignments)  
1904.368 Million cell updates/sec

Title: US-09-689-952-2

Perfect score: 1643  
Sequence: 1 MCGGSGIMKQFKSIINTSQD.....ERVKSLSTPYPLSGENFRNN 313

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database:

Published Applications AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCT05\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	110	6.7	866	US-09-841-260-113	Sequence 113, App
2	110	6.7	866	US-10-007-693-113	Sequence 113, App
3	105	6.4	248	US-10-287-274-435	Sequence 435, App
4	102	6.2	389	US-10-128-714-3369	Sequence 3369, App
5	102	6.2	389	US-10-128-714-3369	Sequence 8369, App
6	101.5	6.2	827	US-10-032-585-7621	Sequence 7621, App
7	101	6.1	451	US-10-128-714-8421	Sequence 8421, App
8	100	6.1	653	US-10-156-761-11850	Sequence 11850, A
9	100.4	6.1	740	US-10-094-749-2991	Sequence 2991, App
10	99.5	6.1	1177	US-10-128-714-3421	Sequence 3421, App
11	98	6.0	393	US-10-128-714-3421	Sequence 3421, App
12	98	6.0	467	US-09-815-242-10395	Sequence 10395, A
13	97.5	5.9	867	US-09-841-260-109	Sequence 109, App
14	97.5	5.9	867	US-10-007-693-109	Sequence 109, App
15	97.5	5.9	1179	US-10-128-714-8493	Sequence 8493, App

16	97	5.9	389	10	US-09-976-165-13	Sequence 13, App1
17	97	5.9	389	12	US-10-205-219-165	Sequence 165, App
18	97	5.9	389	12	US-10-342-276-13	Sequence 13, App1
19	96.5	5.9	442	9	US-09-815-242-13552	Sequence 13552, A
20	96.5	5.9	652	9	US-09-351-794A-2	Sequence 7047, App
21	96.5	5.9	652	12	US-10-032-585-7047	Sequence 11167, A
22	96	5.8	454	9	US-09-815-242-11167	Sequence 7544, App
23	96	5.8	1368	12	US-10-032-585-7544	Sequence 5010, App
24	95	5.8	445	9	US-09-815-242-5010	Sequence 10908, A
25	95	5.8	447	9	US-09-815-242-10908	Sequence 7284, App
26	94.5	5.8	411	12	US-10-032-585-7284	Sequence 7201, App
27	94.5	5.8	796	12	US-10-032-585-7201	Sequence 51, App1
28	94.5	5.8	942	12	US-10-334-703-51	Sequence 51, App1
29	94.5	5.8	942	12	US-10-396-122-51	Sequence 10193, A
30	93.5	5.7	577	9	US-09-815-242-10193	Sequence 91, App1
31	93.5	5.7	1576	12	US-10-336-122-91	Sequence 99, App1
32	93.5	5.7	1940	12	US-09-738-630-99	Sequence 984, App
33	93	5.7	402	9	US-09-925-301-984	Sequence 1774, App
34	92.5	5.6	574	10	US-09-764-868-774	Sequence 12048, A
35	92	5.6	870	15	US-10-156-761-12048	Sequence 20, App1
36	91.5	5.6	284	9	US-09-861-451A-20	Sequence 6009, App
37	91.5	5.6	350	10	US-09-738-626-6009	Sequence 7325, App
38	91.5	5.6	401	12	US-10-032-585-7325	Sequence 7319, App
39	91.5	5.6	1240	12	US-10-032-585-7319	Sequence 29, App1
40	91	5.5	2125	10	US-09-919-172-29	Sequence 169, App1
41	91	5.5	2649	12	US-10-205-219-169	Sequence 220, App
42	91	5.5	2649	12	US-10-341-434-220	Sequence 230, App
43	91	5.5	2649	12	US-10-341-434-230	Sequence 43043, A
44	90.5	5.5	304	9	US-09-864-761-43043	Sequence 102, App
45	90.5	5.5	508	11	US-09-769-787-102	

#### ALIGNMENTS

RESULT 1  
US-09-841-260-113  
Sequence 113, Application US/09841260  
Publication No. US20030175700A1  
GENERAL INFORMATION:  
APPLICANT: Bhatia, Ajay  
APPLICANT: Probst, Peter  
APPLICANT: Stromberg, Erika Jean  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND DIAGNOSIS  
FILE REFERENCE: 210121.515  
CURRENT APPLICATION NUMBER: US/09/841,260  
CURRENT FILING DATE: 2001-04-23  
NUMBER OF SEQ ID NOS: 140  
SEQ ID NO 113  
LENGTH: 866  
TYPE: PRT  
ORGANISM: Chlamydia pneumoniae  
US-09-841-260-113

Query Match 6.7%, Score 110, DB 12, Length 866;  
Best Local Similarity 22.0%; Pred. No. 0.12;  
Matches 81, Conservative 71, Mismatches 123, Indels 94, Gaps 20;

QY	DB	QY	DB	QY	DB	QY	DB	QY	DB
7	IKKQPKSIINTSQDEKRIKIKKVINDDPVKQFLKARALTNAMI--DDLNLVLOEY 64	422	IKKQAIKREQSPSYOEADMQKSI-----DLRELALSLRLGMDKELISL 471						
65	KDQQRHYDGHKFCADPNFKVGHVPELVYDNNRI-KIRYQCC--KIKYD----- 112	472	KEKKNSLSMKFS-----EAEARVADYNNVALRYSLIPQLREIKQDASLNRQDN 524						
113	-----ERFAELI-----TSHMOR-----DTLAKIKDLYMNRDL-----DVAMAADI 154	525	RLQREVDRLIAQVAVNMGTGIPVQKMLGEAKELLLIESLLEKRVVGQPFVAVSDSI 584						
155	CTA---ITNGEOKVLVLY-GPFGTKSFTILGAINOLKSKKVRSTIYLPRF-----IR 205								

Db 585 RAARVGLNDPQRLVFLPGLPTGKTEKLAKALADLLFNKEBAMVRFDMSEYMEKHSIS 644  
Qy 206 TLNG-----GFKGSGPEKHLHREANILMLDIDGAEVTPWVRDEYIGPLHY----- 254  
Db 645 KLIGSSPGYGYEGGSLSEALRRRPSVYLFDEIRAD-----KEVLNILLQVFPDGI 698  
Qy 255 -----RMVHELPTFF--SSNFDYSELEHHLAMTRDDEEKTAKARIETRVKSLSTPY--- 303  
Db 699 LTGDKRRKVCNKALFMTSNIGSPELADY--CSKKSSELTK-----EALISVSPVLKR 751  
Qy 304 FLGGENFRN 312  
Db 752 YLSPE-FMN 759

RESULT 2  
US-10-007-693-113  
; Sequence 113, Application US/10007693  
; Publication No. US2002014676A1  
; GENERAL INFORMATION:  
; APPLICANT: Bhatia, Ajay  
; APPLICANT: Probec, Fecet  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT  
; TITLE OF INVENTION: AND DIAGNOSIS OF CHLAMYDIAL INFECTION  
; FILE REFERENCE: 210121.515C2  
; CURRENT APPLICATION NUMBER: US/10/007,693  
; NUMBER OF SEQ ID NOS: 157  
; SEQ ID NO 113  
; LENGTH: 866  
; TYPE: PRT  
; ORGANISM: Chlamydia pneumoniae  
US-10-007-693-113

Query Match 6.7%; Score 110; DB 14; Length 866;  
Best Local Similarity 22.0%; Pred. No. 0.12;  
Matches 81; Conservative 71; Mismatches 123; Indels 94; Gaps 20;  
Qy 7 IMPOKSIINTSODPEKRIEKIKKEVINDDPVQPLFANRAELTNMI--DEDLNVLOEY 64  
Db 422 IVKQEAIRKQSPSYQEAADAMQKSI-----DALREELASLRIGWDEERKQLISGL 471  
Qy 65 KDOOKHYDGHKAFADCPHFVKGHVVELYVNNRI-KIRYLOQPC--KIKYDE----- 112  
Db 472 KEKKNSJESKFS-----EEHRYADYRVAELKYSILPOLBEETKQDEHSLNGRDN 524  
Qy 113 -----ERPEALI--TSHMQR--DTLNALKDIYNNHRDL-----DYMAADDI 154  
Db 525 RLQGEVDERLIQVAVNMWTGIPVQKMLEGAELKLLIEESLEERVVGGPFAVSAVDSI 584  
Qy 155 CTA---ITNGEQVGLVY-GPFGTGSFLIGAIANOLKSKVSTIILPEF-----IR 205  
Db 585 RAARVGLNDPQRLVFLPGLPTGKTEKLAKALADLLFNKEBAMVRFDMSEYMEKHSIS 644  
Qy 206 TLNG-----GFKDSEFEKHLHREANILMLDIDGAEVTPWVRDEYIGPLHY----- 254  
Db 645 KLIGSSPGYGYEGGSLSEALRRRPSVYLFDEIRAD-----KEVLNILLQVFPDGI 698  
Qy 255 -----RMVHELPTFF--SSNFDYSELEHHLAMTRDDEEKTAKARIETRVKSLSTPY--- 303  
Db 699 LTGDKRRKVCNKALFMTSNIGSPELADY--CSKKSSELTK-----EALISVSPVLKR 751  
Qy 304 FLGGENFRN 312  
Db 752 YLSPE-FMN 759

RESULT 3  
US-10-287-274-435  
; Sequence 435, Application US/10287274  
; Publication No. US20030181408A1  
; GENERAL INFORMATION:  
; APPLICANT: Foreyth, R. Allyn

; APPLICANT: Ohlsen, Karl  
; APPLICANT: Zykkind, Judith  
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERETO  
; FILE REFERENCE: ELITTA.008DV1  
; CURRENT APPLICATION NUMBER: US/10/287,274  
; PRIOR FILING DATE: 2002-10-31  
; PRIOR APPLICATION NUMBER: US 60/164415  
; PRIOR FILING DATE: 1999-11-09  
; PRIOR APPLICATION NUMBER: US 09/711164  
; PRIOR FILING DATE: 2000-11-09  
; NUMBER OF SEQ ID NOS: 469  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 435  
; LENGTH: 248  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-10-287-274-435

Query Match 6.4%; Score 105; DB 12; Length 248;  
Best Local Similarity 26.7%; Pred. No. 0.062;  
Matches 43; Conservative 30; Mismatches 66; Indels 22; Gaps 7;  
Qy 140 NHRDLVANA--ADICTAINGQVKGLYYGPFGTGSFLIGAIANOLSKKVRST 196  
Db 80 NEGQVYALSQAKSIADBLMTGCTN-----FAFSKRGTKKMLAAIGRLKDGQTVI 133  
Qy 197 IYLPFRIRTKGFRDG-SFEKKLRVREANILMLDIDGAEVTPWVRDEYIGPLHYR 255  
Db 134 VVTVADVMSALHASYDGGSGEKLRELCVDLVLVDEIGIQTET--KNEQV--VLH-Q 187  
Qy 256 MVHELPTFFSSNFDYSELEHHLAMTRDDEEKTAKARIETRV 296  
Db 188 IYDRRTASMSVGMLTNINYEAMKTLIGR-----RIMDMN 222

RESULT 4  
US-10-128-714-3369  
; Sequence 3369, Application US/10128714  
; Publication No. US20030119013A1  
; GENERAL INFORMATION:  
; APPLICANT: Jiang, Bo  
; APPLICANT: Hu, Wengdi  
; APPLICANT: Fishkoef, Daniel  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Eroshkin, Alexey M  
; APPLICANT: Lemieux, Sebastien M  
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and  
; FILE REFERENCE: 10182-018-999  
; CURRENT APPLICATION NUMBER: US/10/128,714  
; PRIOR FILING DATE: 2002-04-23  
; PRIOR APPLICATION NUMBER: US 60/285,697  
; PRIOR FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: US 60/287,066  
; PRIOR FILING DATE: 2001-04-27  
; PRIOR APPLICATION NUMBER: US 60/295,890  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US 60/303,899  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: US 60/316,362  
; PRIOR FILING DATE: 2001-08-31  
; NUMBER OF SEQ ID NOS: 8603  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 3369  
; LENGTH: 389  
; TYPE: PRT  
; ORGANISM: Aspergillus fumigatus  
US-10-128-714-3369

Query Match 6.2%; Score 102; DB 15; Length 389;  
Best Local Similarity 21.5%; Pred. No. 0.23;  
Matches 67; Conservative 60; Mismatches 104; Indels 80; Gaps 15;

```

QY 21 FEKRIKIKKEVINDPVKQFLAHRALTN--AMIDEDINVLQOEYKQOKHYDGH---K 75
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7 YRKIKESMKLEIIQGAVALRLERAGRDVNSRVRLRLREELGLQ---QCGSYGGEVVKV 62
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 76 FADCPNFVKGHVELYVDN--NRIKIRYLOCPCKIKYDEERFEAE-----LITSHH 124
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 63 MSYKRVLVKHPGCKYVVDIADGVDITKLVGKRVALLSDSYLKEKMLPSSVDPVLISLM 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 125 MOR-----DTLNAKLKDITMNRHRLDVMAADDICTALTNGEQVGLVYGPFGT 175
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 123 VEKVPDSTYDMIGLDOQIKEL---KEVIELGKHPLEFESL-GIAQPGVLLYGPFGT 177
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 176 GKSPFIIGAIANOLSKSKVR---STII--YLPEFIRTLKGFKDSFEKTLHVRANITLM 230
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 178 GKTLARAVAHHTDCKPFRVSGSELVQYIGBSRWRLRF-----VWAREHAPSIIIP 230
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 231 LDDIGAEVTPWVRDEVIGPLHYRMVHELPTFPSSNFDY-----SELEH---HLAMTRD 282
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 231 MDEI-----DSIG-----SSRIDSGSGDSSEVQRTMELLNQD 264
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 283 GEEKTKAARI 293
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 265 GFEPTNKIKI 275
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

# RESULT 5

```

US-10-128-714-8369
/ Sequence 8369, Application US/10128714
/ Publication No. US20030119013A1
/ GENERAL INFORMATION:
/ APPLICANT: Jiang, Bo
/ APPLICANT: Hu, Weng
/ APPLICANT: Tishkoff, Daniel
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Broshkin, Alexey M
/ APPLICANT: Lemieux, Sebastien M
/ TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
/ FILE REFERENCE: 10182-018-999
/ CURRENT APPLICATION NUMBER: US/10/128-714
/ CURRENT FILING DATE: 2002-04-23
/ PRIOR APPLICATION NUMBER: US 60/285,697
/ PRIOR FILING DATE: 2001-04-23
/ PRIOR APPLICATION NUMBER: US 60/287,066
/ PRIOR FILING DATE: 2001-04-27
/ PRIOR APPLICATION NUMBER: US 60/295,890
/ PRIOR FILING DATE: 2001-06-05
/ PRIOR APPLICATION NUMBER: US 60/303,899
/ PRIOR FILING DATE: 2001-07-09
/ PRIOR APPLICATION NUMBER: US 60/316,362
/ PRIOR FILING DATE: 2001-08-31
/ NUMBER OF SEQ ID NOS: 8603
/ SOFTWARE: PatentIn version 3.1
/ LENGTH: 389
/ TYPE: PRT
/ ORGANISM: Aspergillus fumigatus
US-10-128-714-8369

```

Query Match 6.2%; Score 102; DB 15; Length 389;  
 Best Local Similarity 21.5%; Pred. No. 0.23;  
 Matches 67; Conservative 60; Mismatches 104; Indels 80; Gaps 15;

```

QY 21 FEKRIKIKKEVINDPVKQFLAHRALTN--AMIDEDINVLQOEYKQOKHYDGH---K 75
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7 YRKIKESMKLEIIQGAVALRLERAGRDVNSRVRLRLREELGLQ---QCGSYGGEVVKV 62
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 76 FADCPNFVKGHVELYVDN--NRIKIRYLOCPCKIKYDEERFEAE-----LITSHH 124
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 63 MSYKRVLVKHPGCKYVVDIADGVDITKLVGKRVALLSDSYLKEKMLPSSVDPVLISLM 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 125 MOR-----DTLNAKLKDITMNRHRLDVMAADDICTALTNGEQVGLVYGPFGT 175
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

DB 123 VEKVPDSTYDMIGLDOQIKEL---KEVIELGKHPLEFESL-GIAQPGVLLYGPFGT 177
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 176 GKSPFIIGAIANOLSKSKVR---STII--YLPEFIRTLKGFKDSFEKTLHVRANITLM 230
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 178 GKTLARAVAHHTDCKPFRVSGSELVQYIGBSRWRLRF-----VWAREHAPSIIIP 230
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 231 LDDIGAEVTPWVRDEVIGPLHYRMVHELPTFPSSNFDY-----SELEH---HLAMTRD 282
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 231 MDEI-----DSIG-----SSRIDSGSGDSSEVQRTMELLNQD 264
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 283 GEEKTKAARI 293
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 265 GFEPTNKIKI 275
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

# RESULT 6

```

US-10-032-585-7621
/ Sequence 7621, Application US/10032585
/ Publication No. US20030180953A1
/ GENERAL INFORMATION:
/ APPLICANT: Terry, Roemer D.
/ APPLICANT: Bo, Jiang
/ APPLICANT: Charles, Boone
/ APPLICANT: Howard, Bussey
/ TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
/ FILE REFERENCE: 10182-005-999
/ CURRENT APPLICATION NUMBER: US/10/032-585
/ CURRENT FILING DATE: 2001-12-20
/ NUMBER OF SEQ ID NOS: 8000
/ SOFTWARE: PatentIn version 3.1
/ LENGTH: 827
/ TYPE: PRT
/ ORGANISM: Candida albicans
US-10-032-585-7621

```

Query Match 6.2%; Score 101.5; DB 12; Length 827;  
 Best Local Similarity 24.2%; Pred. No. 0.74;  
 Matches 71; Conservative 41; Mismatches 96; Indels 85; Gaps 17;

```

QY 19 ODPE-KRIEK-1KK-----EVINDPDKQFLAHRALTNAMDED-----LNLVL 61
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 64 KDFOLKRMKNKNTLEKSIASLKVIRQDETEBF-----GDRNGIDGSESEANDRECENDL 118
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 62 QRYKD-----QKHYDGHKFADEPNFVKH---VELYVDNR-----IKIRY 101
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 119 MEYKDSNAINKSVSLMNQKTDGE--SDGVNNEBQADTTELKGNRRKAKGSAKTOL 176
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 102 LOCPCKIKYDEERFE-----AELTSHMORDTLNAKLKDITMNRHRLDVMAADDIC 155
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 177 KQOKRKIDYSTNIDSLSGVSEVTTQLLEITGLPILHPEIYSS----- 221
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 156 TAITNGEQVGLVYGPFGTKSFIIGAIANOLSKSKVSTIIYLPFIRTLKGFKDS 215
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 222 ---TGVEPRGVLLVGPFCGKTTIAMLAGELVKPFIN---ISAPSVSGM-----SGE 270
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 216 FEKGLHRV-REAN-----TLMDDIGAEVTPWVRDEVIGPLHYRMVHELPT 262
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 271 SEKULREIFEAKQIAFCLIFMDEIDA--ITD-KRDGGAQREMEKRIYVALULT 320
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

# RESULT 7

```

US-10-128-714-8421
/ Sequence 8421, Application US/10128714
/ Publication No. US20030119013A1
/ GENERAL INFORMATION:
/ APPLICANT: Jiang, Bo
/ APPLICANT: Tishkoff, Daniel
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Broshkin, Alexey M
/ APPLICANT: Lemieux, Sebastien M
/ TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and

```

TITLE OF INVENTION: Methods of Use  
FILE REFERENCE: 10182-018-999  
CURRENT APPLICATION NUMBER: US/10/128,714  
CURRENT FILING DATE: 2002-04-23  
PRIOR APPLICATION NUMBER: US 60/285,697  
PRIOR FILING DATE: 2001-04-23  
PRIOR APPLICATION NUMBER: US 60/287,066  
PRIOR FILING DATE: 2001-04-27  
PRIOR APPLICATION NUMBER: US 60/295,890  
PRIOR FILING DATE: 2001-06-05  
PRIOR APPLICATION NUMBER: US 60/303,899  
PRIOR FILING DATE: 2001-07-09  
PRIOR APPLICATION NUMBER: US 60/316,362  
PRIOR FILING DATE: 2001-08-31  
NUMBER OF SEQ ID NOS: 8603  
SOFTWARE: Patent version 3.1  
SEQ ID NO 8421  
LENGTH: 451  
TYPE: PRT  
ORGANISM: *Aspergillus fumigatus*  
US-10-128-714-8421

Query Match 6.1%; Score 101; DB 15; Length 451;  
Best Local Similarity 20.8%; Pred. No. 0.35;  
Matches 52; Conservative 51; Mismatches 87; Indels 60; Gaps 12;

QY 34 NDDPVKQFLAHRAELTNAMIDED-LNVQEVYDQOKHYD---GHKPADCPYVYGHV 87  
DB 63 SDPEREQALDYKSLLELEWEKAKNLNMGIDQREPDISSENTKALQSVQIIGEV 122  
QY 88 PELVYDNNRIKIRYLQCP---CKIYDEERFE-----AELITSHMORDT----- 129  
DB 123 LK-QLBERRYIVASSSPRYVGRSKVDKSKQGRVALDMTTLTMMLEPREVDPLV 181  
QY 130 -----LNAKLDIYMHNRDLVMAADICTALTNGEYVQKLYLYG 171  
DB 182 YNMSLEDPGQISFAGISGLNDQIREL---REVIELELPKPELPQRY-GIKPPKGVLLYG 236  
QY 172 PRTGKSFILGAIANOLKS---KKYRSTII--YIPETIRLKGFKGSEKULHREA 226  
DB 237 PPGTKTLARAAVASMETFLKVASAIYDKYIGESARLIREWF---GYAKE---HEP 289  
QY 227 NILLMDIGA 236  
DB 290 CIIFMEIDA 299

## RESULT 8

US-10-156-761-11850  
Sequence 11850, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SAKAKI, YOSHITAKA  
APPLICANT: SAKAKI, YOSHITAKA  
APPLICANT: HATTORI, MASAHITO  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 11850  
LENGTH: 653  
TYPE: PRT  
ORGANISM: *Streptomyces avermitilis*  
US-10-156-761-11850

Query Match 6.1%; Score 100; DB 15; Length 653;  
Best Local Similarity 39.2%; Pred. No. 0.74;  
Matches 29; Conservative 15; Mismatches 20; Indels 10; Gaps 4;

QY 167 LVLVYGPFGTKSFILGAIANOLKS---KKVR--STIIVLEPIRTLKGFKGSEKUL 220  
DB 350 LFIYSSGGLCKTLHLAHIGHVANSILVGRVRYVSSSEPFNEFINSLRDS-KDSDSPK-- 406  
QY 221 HRVREANITLMDI 234  
DB 407 -RYREMDIILVDDI 419

## RESULT 9

US-10-094-749-2991  
Sequence 2991, Application US/10094749  
Publication No. US20030219741A1  
GENERAL INFORMATION:  
APPLICANT: ISOGAI, TAKAO  
APPLICANT: SUGIYAMA, TOMOYASU  
APPLICANT: OTSUKI, TETSUJI  
APPLICANT: WAKAMATSU, AI  
APPLICANT: SATO, HIROYUKI  
APPLICANT: ISHII, SHIZUKO  
APPLICANT: YAMAMOTO, JUN-ICHI  
APPLICANT: ISONO, YUTAKO  
APPLICANT: HIO, YUKI  
APPLICANT: OTSUKA, KAORU  
APPLICANT: NAGAI, KEIICHI  
APPLICANT: IRIE, RYOTARO  
APPLICANT: TAMECHIKA, ICHIRO  
APPLICANT: SEKI, NAOKIKO  
APPLICANT: YOSHIKAWA, TSUTOMU  
APPLICANT: OTSUKA, MOTOKYUKI  
APPLICANT: NAGAHARI, KENJI  
APPLICANT: MASUHO, YASUHIKO  
TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA  
FILE REFERENCE: 084335/0160  
CURRENT APPLICATION NUMBER: US/10/094,749  
CURRENT FILING DATE: 2002-03-12  
PRIOR APPLICATION NUMBER: 60/350,435  
PRIOR FILING DATE: 2002-01-24  
PRIOR APPLICATION NUMBER: JP 2001-328381  
PRIOR FILING DATE: 2001-09-14  
NUMBER OF SEQ ID NOS: 3381  
SOFTWARE: Patent Ver. 2.1  
SEQ ID NO 2991  
LENGTH: 740  
TYPE: PRT  
ORGANISM: *Homo sapiens*  
US-10-094-749-2991

Query Match 6.1%; Score 100; DB 12; Length 740;  
Best Local Similarity 21.3%; Pred. No. 0.88;  
Matches 61; Conservative 50; Mismatches 97; Indels 78; Gaps 14;

QY 5 QGIMKQPKSTINTSQPFRIKIKKEVI--NDPVKQFLAHRA--ELTNAMIDEDLN 59  
DB 212 KNVMKTFREEL---VNIKAFEVERQELASNCKKKEQALQANNAELETLN-----N 261  
QY 60 VLQEVYDQOKHYDGHFADCPNFVKGHVPELYDNNRIKIRYLQCPCKIKYDEERFEAL 119  
DB 262 RMKKVDYERKQALRQRIIMDCERY-----NMKTI-----KLEQDVQILBQL 302  
QY 120 ITSHHQDPTLNAKLDIYMHNRDL---YMAADICTALTNGEYVQKLYYGPFGT 175  
DB 303 -----QQR-----KATYQLONEKLEYNLQVTKRDEESTVYKSGQKKRINLHD--- 346  
QY 176 GKSFILGAIANOLKSKRVSTIIVLEPIRTLKGFKGSEKULHREA 225  
DB 347 -----ILNNLRSKYAKQ-IKQFQENQSLTSYKRLVMQFKELQKAMRHFALIDDEK 397

QY 236 ABEVTPVDEVIPLHYRMWELPTFSSNEDVSEL--EHLAM 279  
 Db 398 FRIE-WIMNE-----BEADLAPADVRIHTHGL 430

## RESULT 10

US-10-128-714-3493  
 ; Sequence 3493, Application US/10128714  
 ; Publication No. US20030119013A1

## GENERAL INFORMATION:

APPLICANT: Jiang, Bo  
 APPLICANT: Hu, Weng  
 APPLICANT: Tishkoff, Daniel  
 APPLICANT: Zamudio, Carlos  
 APPLICANT: Eroshkin, Alexey M  
 APPLICANT: Lemieux, Sebastien M  
 TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and  
 FILE REFERENCE: 10182-018-999  
 CURRENT APPLICATION NUMBER: US/10/128,714  
 CURRENT FILING DATE: 2002-04-23  
 PRIOR APPLICATION NUMBER: US 60/285,697  
 PRIOR FILING DATE: 2001-04-23  
 PRIOR APPLICATION NUMBER: US 60/287,066  
 PRIOR FILING DATE: 2001-04-27  
 PRIOR APPLICATION NUMBER: US 60/295,890  
 PRIOR FILING DATE: 2001-06-05  
 PRIOR APPLICATION NUMBER: US 60/303,899  
 PRIOR FILING DATE: 2001-07-09  
 PRIOR APPLICATION NUMBER: US 60/316,362  
 PRIOR FILING DATE: 2001-08-31  
 NUMBER OF SEQ ID NOS: 8603  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO 3493  
 LENGTH: 1177  
 TYPE: PRF  
 ORGANISM: Aspergillus fumigatus  
 US-10-128-714-3493

Query Match 6.1%; Score 99.5; DB 15; Length 1177;  
 Best Local Similarity 25.5%; Pred. No. 1.9;  
 Matches 64; Conservative 33; Mismatches 89; Indels 65; Gaps 13;

QY 13 SIINTSODEKRIEIKKIVINPDVKQF-----LEHRAELTNMIDELNLTQEX 64  
 Db 742 SIIQAVEEMKSNIEOLKQK-IDDAKTRQAEAMKDIRIEKMSSEFNKK-DSKLAELQAT 799  
 QY 65 KQOQKHVDGKFPADCFNFGVPELYVDNNRIKIRYLOCPCKIKYDEERF-EAELITSH 123  
 Db 800 LBSLK-----KGLAKNSNSYKALQKELQIS-----RLSEQVGSLSAAEQYALAE----- 846  
 QY 124 HMQRDTLNAKLDIYNNHRDLVDVMAADICTAINTGEQVKGLYVGPFGTKSPILGA 183  
 Db 847 -----NLIKQMEBIEQMKREO---ADDAHDIAQAQLEDEBAK-----LTGFDELEB 891  
 QY 184 IANQLSKKVRSTIITLPEFIRLTKGFGDGSFEKGL-HRVRAANILMDDIGA----- 236  
 Db 892 LERAISSKNSRLT-----BERGLEMKLGHQLEK---LQKQQAATAQVAH 933  
 QY 237 -BEVTPVDE 246  
 Db 934 MEHEHEMIAD 944

## RESULT 11

US-10-128-714-3421  
 ; Sequence 3421, Application US/10128714  
 ; Publication No. US20030119013A1

## GENERAL INFORMATION:

APPLICANT: Jiang, Bo  
 APPLICANT: Hu, Weng  
 APPLICANT: Tishkoff, Daniel  
 APPLICANT: Zamudio, Carlos

APPLICANT: Eroshkin, Alexey M  
 APPLICANT: Lemieux, Sebastien M  
 TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and  
 FILE REFERENCE: 10182-018-999  
 CURRENT APPLICATION NUMBER: US/10/128,714  
 CURRENT FILING DATE: 2002-04-23  
 PRIOR APPLICATION NUMBER: US 60/285,697  
 PRIOR FILING DATE: 2001-04-23  
 PRIOR APPLICATION NUMBER: US 60/287,066  
 PRIOR FILING DATE: 2001-04-27  
 PRIOR APPLICATION NUMBER: US 60/295,890  
 PRIOR FILING DATE: 2001-06-05  
 PRIOR APPLICATION NUMBER: US 60/303,899  
 PRIOR FILING DATE: 2001-07-09  
 PRIOR APPLICATION NUMBER: US 60/316,362  
 PRIOR FILING DATE: 2001-08-31  
 NUMBER OF SEQ ID NOS: 8603  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO 3421  
 LENGTH: 393  
 TYPE: PRF  
 ORGANISM: Aspergillus fumigatus  
 US-10-128-714-3421

Query Match 6.0%; Score 98; DB 15; Length 393;  
 Best Local Similarity 20.8%; Pred. No. 0.56;  
 Matches 52; Conservative 50; Mismatches 88; Indels 60; Gaps 12;

QY 34 NDPVKQFLEHRAELTNMIDED--LNLQEXKQOQKHVD---GHKPADCPNFVKGHV 87  
 Db 5 SDPERQALDEDYKSLBLERWEAKLNRMGIKOLQRFDISSENIKALQSVGILGV 64  
 QY 88 PELVYDNNRIKIRYQCP-----CKIKYDEERF-----ABLITSHMQRDT---- 129  
 Db 65 LK-QLDEREFIVKASGRYVVGCRSKYDKSKLQKQTRVALDMTTLITIRMLPREVDPLV 123  
 QY 130 -----LNALKLDIYNNHRDLVDVMAADICTAINTGEQVKGLYLG 171  
 Db 124 YNMSLEDEQGISFAGIGLMDQIREL---REVIEPLKNPFLQRV-GIKPPKGVLLYG 178  
 QY 172 PRGTGKSFILGAIANOLKS---KVRSTII--YLPEFIRLTKGFGDGSFEKGLHVRBA 226  
 Db 179 PRGTGKTLARAVASMEINFLKVVSSALVDKIGESARLIREMF---GYAE---HFP 231  
 QY 227 NIMLMDDIGA 236  
 Db 232 CIIFMDEIDA 241

## RESULT 12

US-09-815-242-10395  
 ; Sequence 10395, Application US/09815242  
 ; Patent No. US20020061569A1

## GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert  
 APPLICANT: Ohlsen, Karl T.  
 APPLICANT: Zyskind, Judith W.  
 APPLICANT: Wall, Daniel  
 APPLICANT: Trawick, John D.  
 APPLICANT: Carr, Grant J.  
 APPLICANT: Yamamoto, Robert T.

TITLE OF INVENTION: Identification of Essential Genes in  
 FILE REFERENCE: ELITRA.011A  
 CURRENT APPLICATION NUMBER: US/09/815,242  
 CURRENT FILING DATE: 2001-03-21  
 PRIOR APPLICATION NUMBER: 60/191,078  
 PRIOR FILING DATE: 2000-03-21  
 PRIOR APPLICATION NUMBER: 60/206,848  
 PRIOR FILING DATE: 2000-05-23  
 PRIOR APPLICATION NUMBER: 60/207,727

```

; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 10395
; LENGTH: 467
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10395

Query Match      6.0%; Score 98; DB 9; Length 467;
Best Local Similarity 27.1%; Pred. No. 0.72;
Matches 19; Conservative 21; Mismatches 28; Indels 2; Gaps 1;

QY 167 LVIYGPRTGKSFILGAIANOLSKKVRSTIYL-PEFRTLKGFKDSFEKGLHVR 224
DB 168 LFIYGGTGLKTHLHVGNGIMAKPKNAKVYHSERFVQDVVKALQNNABEEFKRYR 227

QY 225 EANIMDDI 234
DB 228 SVDALLDI 237

RESULT 13
US-09-841-260-109
; Sequence 109, Application US/09841260
; Publication No. US20030175700A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Probst, Peter
; APPLICANT: Stromberg, Erika Jean
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND DIAGNOSIS
; TITLE OF INVENTION: OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.515
; CURRENT APPLICATION NUMBER: US/09/841,260
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 140
; SEQ ID NO 109
; LENGTH: 867
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis serovar D
US-09-841-260-109

Query Match      5.9%; Score 97.5; DB 12; Length 867;
Best Local Similarity 19.9%; Pred. No. 1.9; Indels 105; Gaps 18;
Matches 71; Conservative 71; Mismatches 110; Indels 105; Gaps 18;

QY 18 SDFEKRIEIKKEVINDPVKQFLNRAELTNAMIDEDLVLOEQYKQKHGKFA 77
DB 440 AEDMQAIDRVKE-----LAALRLRW-----DEKGLITGKEKNLENLKFA 484

QY 78 DCNPFVKGHVPELYVNNRI-KIRYLQPC-----KITYDE--ERFEA 117
DB 485 E-----BEAERTADYNRAELRYSILPSLEERHILAEALNQRDGRLLQEBVEDRLIA 537

QY 118 ELI---TSHMOR--DTLNALNDIYNNHRDL-----DVAMAADDICTA---ITNGEOV 164
DB 538 QVAAWMTGIVQKMLEGESEKLVLESLEERVVGCPALAAVSDSIRARVGLSDPQR 597

QY 165 KGLIYV-GPFGTKSFILGAIANOLSKKVRSTIYLPF-----ITLKG-----GFK 212
DB 598 LGVFLFGPTGVGKTELAKALABELFNKEBAMIRFDMTEWMEKGSVKLIGSPGYGYE 657

QY 213 DGSFEKGLHVRANIMDDIGAEEVTPVNRDEVIGPLIHY-----RMHELP 261
DB 658 EGGSLSEALRRRPRYSVLPDEIERKAD-----KEVNIILQIFDDGILTDSKKRKVCN 711
```

```

QY 262 TFF--SSNFDYSELHHLAMTRDSEKTKAARIER-----VKSISTEYFLS 306
DB 712 ALFMTSNIGSOBLADYC-----TKKGTIVDKAVALSVAPALKNYFSSEFIN 759

RESULT 14
US-10-007-693-109
; Sequence 109, Application US/10007693
; Publication No. US20020146776A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Probst, Peter
; APPLICANT: Stromberg, Erika Jean
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT
; TITLE OF INVENTION: AND DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.515C2
; CURRENT APPLICATION NUMBER: US/10/007,693
; CURRENT FILING DATE: 2001-12-05
; NUMBER OF SEQ ID NOS: 157
; SEQ ID NO 109
; LENGTH: 867
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis serovar D
US-10-007-693-109

Query Match      5.9%; Score 97.5; DB 14; Length 867;
Best Local Similarity 19.9%; Pred. No. 1.9; Indels 105; Gaps 18;
Matches 71; Conservative 71; Mismatches 110; Indels 105; Gaps 18;

QY 18 SDFEKRIEIKKEVINDPVKQFLNRAELTNAMIDEDLVLOEQYKQKHGKFA 77
DB 440 AEDMQAIDRVKE-----LAALRLRW-----DEKGLITGKEKNLENLKFA 484

QY 78 DCNPFVKGHVPELYVNNRI-KIRYLQPC-----KITYDE--ERFEA 117
DB 485 E-----BEAERTADYNRAELRYSILPSLEERHILAEALNQRDGRLLQEBVEDRLIA 537

QY 118 ELI---TSHMOR--DTLNALNDIYNNHRDL-----DVAMAADDICTA---ITNGEOV 164
DB 538 QVAAWMTGIVQKMLEGESEKLVLESLEERVVGCPALAAVSDSIRARVGLSDPQR 597

QY 165 KGLIYV-GPFGTKSFILGAIANOLSKKVRSTIYLPF-----ITLKG-----GFK 212
DB 598 LGVFLFGPTGVGKTELAKALABELFNKEBAMIRFDMTEWMEKGSVKLIGSPGYGYE 657

QY 213 DGSFEKGLHVRANIMDDIGAEEVTPVNRDEVIGPLIHY-----RMHELP 261
DB 658 EGGSLSEALRRRPRYSVLPDEIERKAD-----KEVNIILQIFDDGILTDSKKRKVCN 711

QY 262 TFF--SSNFDYSELHHLAMTRDSEKTKAARIER-----VKSISTEYFLS 306
DB 712 ALFMTSNIGSOBLADYC-----TKKGTIVDKAVALSVAPALKNYFSSEFIN 759

RESULT 15
US-10-128-714-8493
; Sequence 8493, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengli
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Broshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
```

;; PRIOR FILING DATE: 2001-04-27  
;; PRIOR APPLICATION NUMBER: US 60/295,890  
;; PRIOR FILING DATE: 2001-06-05  
;; PRIOR APPLICATION NUMBER: US 60/303,899  
;; PRIOR FILING DATE: 2001-07-09  
;; PRIOR APPLICATION NUMBER: US 60/316,362  
;; PRIOR FILING DATE: 2001-08-31  
;; NUMBER OF SEQ ID NOS: 8603  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 8493  
;; LENGTH: 1179  
;; TYPE: PRT  
;; ORGANISM: Aspergillus fumigatus  
US-10-128-714-8493

Query Match 5.9%; Score 97.5; DB 15; Length 1179;  
Best Local Similarity 25.5%; Pred. No. 3;  
Matches 64; Conservative 33; Mismatches 91; Indels 63; Gaps 13;  
QY 13 SIINTSODEFEKRIEIKKEVINPDVQF-----LEAHRAELTYAMIDEDLVNLOEY 64  
DB 742 SIIOAVEBKMSNTEQLKQD-ISDAKTRQAEAMQDIKIEKQMSFNNK-DSKLAELQAT 799  
QY 65 KDOQKHVDGKHFADCPNFVKGVPELVYDNNRIKIRYLQCPCKIKYDERF-EABLITSH 123  
DB 800 LBSLK-----KKLAKNSNSVVALQKELQIS----RLSEBQVGSPLSAEBQYAEAE----- 846  
QY 124 HMRDITLAKLDIYVNHRRDLVYMAAADICTAITNGEYVKGILYGPFGTGSFTLGA 183  
DB 847 ---NTLKQWMEIIOGKREQARV-KDAHDIAQOLEDERAK-----LTGFDEBLRE 893  
QY 184 IANQLSKKVRSTIYLPERITRLKGGFKDGSFEKKL-HRVREANILMLDDIGA----- 236  
DB 894 LEAATSKSKSRIT-----EEGLEMKLGHOLEK--LQKDOQAAQTVAH 935  
QY 237 -EEVTPWVRDE 246  
DB 936 MEEHEHMIADDE 946

Search completed: December 15, 2003, 15:19:02  
Job time : 31.5681 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 15:11:40 ; Search time 17.652 Seconds  
(Without alignments)  
1705.235 Million cell updates/sec

Title: US-09-689-952-2

Perfect score: 1643

Sequence: 1 MGCGSIRMKQFSIINTSD.....ERVKSLSTPYFLSGENFRNN 313

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Database :

1: PIR 76:1\*  
2: p1r2:1\*  
3: p1r3:1\*  
4: p1r4:1\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1607	97.8	306	2	A89952
2	1631	38.4	311	1	I08544
3	601.5	36.6	311	2	H84042
4	552.5	33.6	307	2	AH1269
5	547.5	33.3	307	2	AB1632
6	407.5	24.8	298	2	B98066
7	403.5	24.6	293	2	H86718
8	392.5	23.9	298	2	D95199
9	173	10.5	209	2	S77882
10	145	8.8	282	2	D97138
11	145	8.8	316	2	B82907
12	138.5	8.4	235	2	E70378
13	137.5	8.4	399	2	G70328
14	133.5	8.1	313	2	D69945
15	131.5	8.0	329	2	D97340
16	128.5	7.8	739	1	A70204
17	127	7.7	266	2	I40411
18	126	7.7	287	2	AD1588
19	126	7.7	287	2	AG1651
20	126	7.7	440	2	D81615
21	120.5	7.3	316	2	T10436
22	117.5	7.2	267	2	E84092
23	117.5	7.2	449	2	A83650
24	117.5	7.2	454	2	D84931
25	116	7.1	842	2	G90576
26	115.5	7.0	246	2	D84932
27	115	7.0	410	2	C69197
28	114	6.9	265	1	BVEIC1
29	112.5	6.8	733	2	H69411

30	112	6.8	291	2	A86681	hypothetical prote
31	112	6.8	520	2	H69125	hypothetical prote
32	111.5	6.8	475	2	C96503	protein P9316.7 (1
33	111	6.8	350	2	B97735	probable ATPase n2
34	111	6.8	1208	2	T39068	coiled coil protei
35	110	6.7	461	2	A90512	chromosomal replic
36	110	6.7	866	2	E72113	clp proteinase ATP
37	110	6.7	866	2	H86508	ATP-dependent Clp
38	110	6.7	872	2	H81556	ATP-dependent Clp
39	109.5	6.7	294	2	C86802	hypothetical prote
40	109	6.6	630	2	AC1309	probable ABC trans
41	109	6.6	791	2	D81310	endopeptidase Ia (
42	109	6.6	831	1	A71825	endopeptidase Ia (
43	108.5	6.6	454	1	I04VBA	replication initia
44	108.5	6.6	759	2	D84301	cell division cycl
45	108	6.6	811	2	B69512	cell division cont

## ALIGNMENTS

RESULT 1  
A89952  
primosomal protein [imported] - Staphylococcus aureus (strain N315)  
C/Species: Staphylococcus aureus  
C/Date: 10-May-2001 #sequence\_revision 10-May-2001 #ext\_change 22-Oct-2001  
C/Accession: A89952  
R/Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi, A.; Mizutani, U.; Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayaishi, H.; Hiratsuku, K.  
A/Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A/Reference number: A89758; PMID:21311952; PMID:11416146  
A/Accession: A89952  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-306 <KUR>  
A/Cross-references: GB:BA000018; PID:g13701480; PIDN:BA42774.1; GSPDB:GN00149  
A/Experimental source: strain N315  
C/Genetics:  
A/Gene: dnaI  
C/Superfamily: 44K dnaI protein homolog

Query Match 97.8%; Score 1607; DB 2; Length 306;  
Best Local Similarity 100.0%; Pred. No. 5.4e-104;  
Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	8	MRQFSIIINTSDPFERIKIKKIVYNDPVKQFLFAHRAELTNAMIDEDLVNLOEYKQD	67
DB	1	MGQFSIIINTSDPFERIKIKKIVYNDPVKQFLFAHRAELTNAMIDEDLVNLOEYKQD	60
QY	68	QKHVDGKHPADCPNVKGVHPELVYDNNRIKIRYLQCPCKIRYDEERFAELITSHMQR	127
DB	61	QKHVDGKHPADCPNVKGVHPELVYDNNRIKIRYLQCPCKIRYDEERFAELITSHMQR	120
QY	128	DTLNKLDIYNNHDDRLDVMAAADICTAINTNGQVGLYIXGPRGKSTILGAINQ	187
DB	121	DTLNKLDIYNNHDDRLDVMAAADICTAINTNGQVGLYIXGPRGKSTILGAINQ	180
QY	188	LKSKVRSITITLPEFIRTLKGFQDGFPEKTLHVRANITIMDDIGAEVTPWVRDEV	247
DB	181	LKSKVRSITITLPEFIRTLKGFQDGFPEKTLHVRANITIMDDIGAEVTPWVRDEV	240
QY	248	IGPLLHYRVNHELPTFFSSNFDYSLEHLAMTRDGEKTKARIIERVKSLSTPYFLSG	307
DB	241	IGPLLHYRVNHELPTFFSSNFDYSLEHLAMTRDGEKTKARIIERVKSLSTPYFLSG	300
QY	308	ENFRNN 313	
DB	301	ENFRNN 306	

RESULT 2

RESULT 3

H84042

primosome component (helicase loader) dnaI [imported] - Bacillus halodurans (strain C-12)

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001

C:Accession: H84042

R:Takami, H.; Nakase, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirano, T.

Nucleic Acids Res. 28, 4317-4331, 2000

A:Article: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A:Reference number: AB3650; MUID:20512582; PMID:11058132

A:Accession: H84042

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-311 <STO>

A:Cross-references: GB:AP001517; GB:BA000004; MID:g10175500; PIDN:BA06863.1; GSPDB:GN001

A:Experimental source: strain C-125

C:Genetics:

A:Gene: dnaI

C:Superfamily: 44k dnaA protein homolog

Query Match 36.6%; Score 601.5; DB 2; Length 311;

Beet Local Similarity 39.2%; Pred. No. 2e-34; Indels 21; Gaps 8;

Matches 125; Conservative 69; Mismatches 104;

QY 5 QSIMKQFSAIINTSODFEKRIEIKKEVINDPVKQFLBNRAELTNAMIDEDINLQY 64

DB 2 ESIQSLKHMTC-NRSFEKQFTQLKEAVRSPHVQFLBHH-PSLSPTILEQSLKLEY 59

QY 65 KDQCKHYDGH-KPADCFPVGVHVELVYDNNRIKIRYIQCECKIKYDERFPAELITS 122

DB 60 QKEQSHC-AHCGLQKCPNLMKGYOPTLYVERSLSTSPCLKEEKEKKRSLRS 118

QY 123 HHMQGDTNANKLKD-----YNNHRDLVDVMAADICTAITNGEQVGLYLYPEGT 175

DB 119 LYIPKEILAKRDVDSERGRSIAHSR-ALFRLSKP-----GPDGMGLYLYGKFGV 170

QY 176 GKSPFLGAIANOLSKKRVASTIYLPEDFRTLKGGFKDSFEKKLHRVREANITMLDDIG 235

DB 171 GKTFLMGAIANELKRGIDSTIVVDPDFRELQSGIDGTFQCKLFFVKAQAVLIPFDIG 230

QY 236 ABEVPMWRDEVICGLHYRWVHELPTFSSNPDVSELEHHLAMT-RDGSEKTPAARLIE 294

DB 231 ABTMTSWRBDVLAGVILQYRIMEKLPFLTSNVDYDELEHHLAYNDKSGTELLKAKRWE 290

QY 295 RVKSLSTPEYFLSGENFRNN 313

DB 291 RIRHYTGSVMVGGNRRH 309

RESULT 4

AHI269

primosome component (helicase loader) dnaI [imported] - Listeria monocytogenes (strain EC

C:Species: Listeria monocytogenes

C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001

C:Accession: AHI269

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, P.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Feihl, H.;

S. Jones, L.M.; Karel, U.

Science 294, 849-852, 2001

A:Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurapkac, G.; Madueno, E.; Maitournam, A.; Mat

ok, C.; Schlueter, T.; Simoes, N.; Tiller, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,

A:Title: Comparative genomics of Listeria species

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AHI269

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-307 <GLA>

A:Cross-references: GB:NC003210; PIDN:CA099638.1; PID:g16410989; GSPDB:GN00177

A:Experimental source: strain EGD-e

C:Genetics:

A:Gene: dnaI

Query Match 33.6%; Score 552.5; DB 2; Length 307;  
 Best Local Similarity 37.8%; Pred. No. 4.7e-31;  
 Matches 113; Conservative 69; Mismatches 104; Indels 13; Gaps 4;

19 QDEKRIEIKKVINDDPVKQFLHRAELTNAMIDEDLNLOEYDQOKHYDGHKPAD 78  
 15 RDEKEKYGKQOVLMHOPIDPFKEKHEVTEQLVNVQNSLNYEFTHQKKEFTEGE--- 71  
 79 CENFVKGHVELYVNNRIRIKIRLQCPCKIKYDEEREAALITSHMQORTLNAKLDIY 138  
 72 -ETLMGAYAKVLNNGEFIVTYVPTKEKIEBKRRRAVERIRISLVPKQVADNADFY 130  
 139 MNHRDLVMAADDICTAITN-----GEQVKGILYVPGPTGKSPFLGAIANOLSKKY 193  
 131 TDESR---QALVAYQPLANTPPKSGERVKGLFIHSGFTGSKYLGLAKELAKGI 187  
 194 RSTIYLPFIRTLKGGFKDSEFKLHVRVANIIMLDIGAEVTPWVRDEVIGPLTH 253  
 188 STLVLPLPFMRVYKQISINTVGEKIQFAKETEVLMLDIGAESMTAMTRDEVILG 247  
 254 YRMVHELPTFFSSNFDYSELHHLAMTRDG-EKTKAARIIEVKSLSTPYPLSGENFR 311  
 248 FRMOEELPTFFSSNFMNDQLENHLMFAQNGTEERLAKARRIMERVRYLSKEVNLGKNRR 306

## RESULT 5

AB1632  
 primosome component (helicase loader) DnaI [imported] - *Listeria innocua* (strain Cl1p112)  
 C:Species: *Listeria innocua*  
 C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
 C:Accession: AB1632  
 R:Glaeser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fehl, H.; Jones, L.M.; Karst, U.  
 Science 294, 849-852, 2001  
 A:Authors: Krefel, J.; Kuhn, M.; Kunst, F.; Kurapkak, G.; Madueno, E.; Maltounam, A.; Maok, C.; Schlueter, T.; Simoes, N.; Tietze, A.; Vazquez-Boland, J.A.; Voss, H.; Weiland, A.; Title: Comparative genomics of *Listeria* species.  
 A:Reference number: AB1077; MUID:21537279; PMID:11679669  
 A:Accession: AB1632  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-307 <GTA>  
 A:Cross-references: GB:AL502022; PIDN:CAC96826.1; PID:G16414082; GSPDB:GN00178  
 A:Experimental source: strain Cl1p11262  
 C:Genetics:  
 A:Gene: dnaI  
 A:Superfamily: 44k dnaA protein homolog

Query Match 33.3%; Score 547.5; DB 2; Length 307;  
 Best Local Similarity 37.8%; Pred. No. 1e-30;  
 Matches 112; Conservative 69; Mismatches 109; Indels 7; Gaps 3;  
 19 ODEKRIEIKKVINDDPVKQFLHRAELTNAMIDEDLNLOEYDQOKHYDGHKPAD 78  
 15 RDEKEKYGKQOVLMHOPIDPFKEKHEVTEQLVNVQNSLNYEFTHQKKEFTEGE--- 71  
 79 CENFVKGHVELYVNNRIRIKIRLQCPCKIKYDEEREAALITSHMQORTLNAKLDIY 138  
 72 -ETLMGAYAKVLNNGEFIVTYVPTKEKIEBKRRRAVERIRISLVPKQVADNADFY 130  
 139 MNHRDLVMAADDICTAI--TNGEYKGLYVPGPTGKSPFLGAIANOLSKKYRST 196  
 131 TDESRDLAVAYQPLANTPPKSGERVKGLFIHSGFTGSKYLGLAKELAKGI 190  
 197 IYLPFIRTLKGGFKDSEFKLHVRVANIIMLDIGAEVTPWVRDEVIGPLTHYR 256  
 191 LVTLPLPFMRVYKQISINTVGEKIQFAKETEVLMLDIGAESMTAMTRDEVILG 250  
 257 VHELPTFFSSNFDYSELHHLAMTRDG-EKTKAARIIEVKSLSTPYPLSGENFR 311  
 251 QSELPTFFSSNFMNDQLENHLMFAQNGTEERLAKARRIMERVRYLSKEVNLGKNRR 306

## RESULT 6

B98066  
 primosome component (helicase loader) [imported] - *Streptococcus pneumoniae* (strain R6)  
 C:Species: *Streptococcus pneumoniae*  
 C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 02-Nov-2001  
 C:Accession: B98066  
 R:Hoshino, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Buggett, S.; Dehoff, B.S.; Ely, P.; Lelanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAnren, S.; Meyer, R.; Sun, P.M.; Winkler, M.B.  
 J. Bacteriol. 183, 5709-5717, 2001  
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaekunas, S.R.; A:Title: Genome of the Bacterium *Streptococcus pneumoniae* Strain R6.  
 A:Reference number: A97872; MUID:21429245; PMID:11544234  
 A:Accession: B98066  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-298 <KUR>  
 A:Cross-references: GB:AE007317; PIDN:AAL00359.1; PID:G15459220; GSPDB:GN00174.  
 C:Genetics:  
 A:Gene: dnaI  
 A:Superfamily: 44k dnaA protein homolog

Query Match 24.8%; Score 407.5; DB 2; Length 298;  
 Best Local Similarity 32.0%; Pred. No. 4.8e-21;  
 Matches 94; Conservative 64; Mismatches 121; Indels 15; Gaps 7;

21 FEKRIEIKKVINDDPVKQFLHRAELTNAMIDEDLNLOEY-KDQGHYDGHKPAD 79  
 17 YDVLQKIMK---DEDAVAFVQ--QESINODELANSISKFNQYITERKFLRG---DT 66  
 80 PNFVKGHVELYVNNRIRIKIRLQCPCKIKYDEEREAALITSHMQORTLNAKLDIY 139  
 67 DYIAKQKPLVNNHGYADVSTETPELIAEKKAIRKRLNINPSSLKVNSTLDVYR 126  
 140 NHRDLVMAADDICTAITNGEYKGLYVPGPTGKSPFLGAIANOLSKK-VSTII 198  
 127 DVQKRLTVLAKRIEFVNDPN--NLKGLVLXGDFGSKGFMVAALAHLSKRGVSTILL 184  
 199 YLPFIRTLKGGFKDSEFKLHVRVANIIMLDIGAEVTPWVRDEVIGPLTHYRVA 258  
 185 HXPSFVIDYKMAISDGNVTVLDEITLSEVLILDDIGAGQSTWVADDEILQVILQYRMOE 244  
 259 ELPTFFSSNFDYSELHHLAMTRDGEK-KAARIIEVKSLSTPYPLSGENFR 311  
 245 NLPTFFSSNFMNDQLENHLMFAQNGTEERLAKARRIMERVRYLSKEVNLGKNRR 298

## RESULT 7

H86718  
 primosomal protein DnaI [imported] - *Lactococcus lactis* subsp. *lactis* (strain IL1403)  
 C:Species: *Lactococcus lactis* subsp. *lactis*  
 C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
 C:Accession: H86718  
 R:Belcetin, A.; Winkler, P.; Mager, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Ehrlich  
 Genome Res. 11, 731-753, 2001  
 A:Title: The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp  
 A:Reference number: A86625; MUID:21335186; PMID:11337471  
 A:Accession: H86718  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-293 <STO>  
 A:Cross-references: GB:AE005176; PID:G12723668; PIDN:AAK04850.1; GSPDB:GN00146  
 A:Experimental source: strain IL1403  
 C:Genetics:  
 A:Gene: dnaI  
 A:Superfamily: 44k dnaA protein homolog

Query Match 24.6%; Score 403.5; DB 2; Length 293;  
 Best Local Similarity 32.3%; Pred. No. 9e-21;  
 Matches 101; Conservative 61; Mismatches 124; Indels 27; Gaps 8;



QY 258 HELPTFSSNPDSYSELEHHLAMTRDGEKTKAARIIRVSKLSPYPLSGENR 311  
 DB 218 SKPLIITSLTLEINPSPKRGVLADQYHERIES-----RIREMCTPVENTSKSIR 266

# RESULT 11

B82907

conserved hypothetical ATP/GTP-binding protein U0107 [imported] - Ureaplasma urealyticum  
 C:Species: Ureaplasma urealyticum  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
 C:Accession: B82907

R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Caswell, G.H.  
 submitted to GenBank, February 2000

A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a mit

A:Reference number: A82870

A:Accession: B82907

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-316 <GAP>

A:Cross-references: GB:AE002128; GB:AF222894; NID:g6899279; PIDN:AAF30716.1; GSPDB:GN001

A:Experimental source: serovar 3; biovar 1

C:Genetics:

A:Gene: U0307

A:Genetic code: SGC3

Query Match 8.8%; Score 145; DB 2; Length 316;  
 Best Local Similarity 21.3%; Pred. No. 0.0075;  
 Matches 69; Conservative 57; Mismatches 130; Indels 68; Gaps 13;

QY 16 NTSODEKRIEIK-KEVINDPVK-----QFLRAHRAELTNAMIDSD-LN 59  
 DB 3 NNOVDDELLENYQKALEIPKAIINTLEEFNLVPMIVAYTSYLNANLQNDSTY 62

QY 60 VLOEYKDOQKHYDGHKPADCPNFVKGHVPELVYDNNRIKIRYLQCPCK-----IKYDE 112  
 DB 63 IYQANNEVYHL-----CLVVVK--KENKVD--KIKKNTIYIANTIKKNDYNISSADDS 110

QY 113 ERR--EAEILTSNM--QRPDLAAKLDIYMRDRDLVMAADICTAITNGEQYKGLY 168  
 DB 111 QIFSKKTEILDANLALERNKLINFLEARKKKNKNTL-----IKNNNQCPKSAF 159

QY 169 LVGPPFGSGFIIGALANOLSKKVRSTIYLPEFRTLLKGPK-----DGSF 216  
 DB 160 IYGDSEVGSIIYQATYNTI-SLKKNKIAYI-----TLMDLFKNVIQFPNYQDTSULVY 213

QY 217 EKKIHRVREANITLMDIGAEVTPVWRDEVIGPLHYRWVHELPTFSSNPDSYSELEH 276  
 DB 214 NELVNLNSVDVVIDDFSSVNLNWSISITLMPITLLENRLKSTRQITFISNFIQGLINS 273

QY 277 LAMTRDGEKTKAARIIRVSKLS 300  
 DB 274 TKNTYIIEBKTKLRLENRIEYLT 297

QY 274 TKNTYIIEBKTKLRLENRIEYLT 297

QY 274 TKNTYIIEBKTKLRLENRIEYLT 297

QY 274 TKNTYIIEBKTKLRLENRIEYLT 297

QY 274 TKNTYIIEBKTKLRLENRIEYLT 297

QY 274 TKNTYIIEBKTKLRLENRIEYLT 297

QY 274 TKNTYIIEBKTKLRLENRIEYLT 297

QY 274 TKNTYIIEBKTKLRLENRIEYLT 297

QY 274 TKNTYIIEBKTKLRLENRIEYLT 297

QY 274 TKNTYIIEBKTKLRLENRIEYLT 297

QY 274 TKNTYIIEBKTKLRLENRIEYLT 297

QY 274 TKNTYIIEBKTKLRLENRIEYLT 297

QY 274 TKNTYIIEBKTKLRLENRIEYLT 297

QY 274 TKNTYIIEBKTKLRLENRIEYLT 297

QY 274 TKNTYIIEBKTKLRLENRIEYLT 297

QY 274 TKNTYIIEBKTKLRLENRIEYLT 297

A:Gene: dnaC

Query Match 8.4%; Score 138.5; DB 2; Length 235;  
 Best Local Similarity 23.8%; Pred. No. 0.014;  
 Matches 59; Conservative 41; Mismatches 93; Indels 55; Gaps 9;

QY 82 FVKGHVPELVYDNNRIKIRYLQCPCKIKYDEERFAELTSHMOBRLNAXKDIYMNH 141  
 DB 15 FVK-----TEDNKRRL-----CECFPKRDVREL-----NIPKRYNANI-DYHPK 56

QY 142 RRLDVMAADICTAITNGEQYKGLYVPGFGTKSFLIGALANOLSKK-VRSITIIYL 200  
 DB 57 NNSQNALITIRVFNPNPBECKGLTFVSGPVGKTHLAVATLKAIVEKKGIRGYFDT 116

QY 201 PERITRLKGGFKDGSFEKTLHVRANITLMDIGAEVTPVWRDEVIGPLHYRWVHEL 260  
 DB 117 KDLIFRLKHMDEGCKTKFRTVNSPVVLDDGSEBSIDWQR-ELISYIITRYRNMLK 175

QY 261 PFPSSNPDSYSELEHHLAMTRDGEKTK-----AARIIRV----- 296  
 DB 176 STITITNY-----SLQREBSVRLSADLASHGENVYSKITYENNELVYKSGDLR 226

QY 297 --KSISTP 302  
 DB 227 KSKKISTP 234

QY 297 --KSISTP 302

QY 297 --KSISTP 302

QY 297 --KSISTP 302

QY 297 --KSISTP 302

QY 297 --KSISTP 302

QY 297 --KSISTP 302

QY 297 --KSISTP 302

QY 297 --KSISTP 302

QY 297 --KSISTP 302

QY 297 --KSISTP 302

QY 297 --KSISTP 302

QY 297 --KSISTP 302

QY 297 --KSISTP 302

QY 297 --KSISTP 302

QY 297 --KSISTP 302

QY 297 --KSISTP 302

QY 297 --KSISTP 302

QY 297 --KSISTP 302

QY 297 --KSISTP 302

QY 297 --KSISTP 302

QY 297 --KSISTP 302

QY 297 --KSISTP 302

QY 297 --KSISTP 302

QY 297 --KSISTP 302

QY 297 --KSISTP 302

QY 297 --KSISTP 302

QY 297 --KSISTP 302

QY 297 --KSISTP 302

C:/Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999

C:/Accession: D69945

C:/Kunet, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berta

C./Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch

A.; Enllich, S.D.; Emerson, P.T.; Enlian, K.D.; Ertington, J.; Fabel, C.; Ferrari, E.

Nature 390, 249-256, 1997

A:/Authors: Foulger, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Gall

leech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holappel, S.; Hosono, S.; Iario, M.F.

Koeter, P.; Koningsveld, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapides, A.; Lardinois,

A./Authors: Lamber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mueel

Y., M.; Ogawa, K.; Ogawa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portecelle

Rieger, M.; Rivolta, C.; Rocha, B.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scallion,

A./Authors: Schleich, S.; Schroeder, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron

akeuchi, M.; Tamkoshi, A.; Tanaka, T.; Terpestra, P.; Togomi, A.; Toesato, V.; Uchiyama,

T.; Winters, P.; Wipet, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K

A./Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A./Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

A./Reference number: A69580; MUID:98044033; PMID:9384377

A./Accession: D69945

A./Status: preliminary; nucleic acid sequence not shown; translation not shown

A./Molecule type: DNA

A./Residues: 1313 <KUN>

A./Cross-references: GB:299117; GB:AL009126; NID:92834966; PIDN:CB14567.1; PID:el183855;

A./Experimental source: strain 168

C/Genetics:

A/Gene: yqam

Query Match 8.1%; Score 133.5; DB 2; Length 313;

Best Local Similarity 24.4%; Pred. No. 0.046;

Matches 66; Conservative 36; Mismatches 90; Indels 83; Gaps 15;

QY 60 VLOEYKQDQKHVYDGHKFA-----DCPNFVKGHVELYVNNRIRYLOCPCKIKYD 111

Db 16 IMGLRERSARFLGTQAVSEKAEKRFDC-----YCKRGIVYRVHKTSTHMD 65

QY 112 EERFPEALITSHM--QRTLNAL-----KDIYNN-----HRDLVAMADIDIC 155

Db 66 E--QDLWVPDDMVSEDDFLGKVCPTPKASEKOTYSKQCECVRRKXIAFLMAASGIT 122

QY 156 -----TAITNG-----EYKQ-----LYLGPFGTGSFLLG 162

Db 123 EEFPEKLFENFIDTGKPDIMKDAVECAVEYKQFQIKGRONSIALLGQSGSKTHLT 162

QY 183 AIANQL-KSKVRSTIYLP--EFIRTLKGGFKDSEFKLHVRREANIIMLDDI----- 234

Db 183 AIAANLIKKSIVH--CMYFVVEGMDLKNF--DNLEAKIDAMRKVEVLFIDDLKPKIN 238

QY 235 GAEVTPWVRDEVIGPLHYRMVHELPTFPSSNFDYSEL 273

Db 239 GQPRATDW-QVEQISVLYNRYLNHKLPLISSELTIDEI 276

RESULT 15

D97340

DNA replication protein DnaC [imported] - Clostridium acetobutylicum

C/Species: Clostridium acetobutylicum

C/Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001

C/Accession: D97340

R/Molling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,

J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A./Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium *Clo*

A./Reference number: A96900; MUID:21359325; PMID:21359325

A./Accession: D97340

A./Status: preliminary

A./Molecule type: DNA

A./Residues: 1-329 <KUR>

A./Cross-references: GB:AE001437; PIDN:AAK8151.1; PID:915026685; GSPDB:GN00168

A./Experimental source: Clostridium acetobutylicum ATCC824

C/Genetics:

A/Gene: CAC3588

Query Match 8.0%; Score 131.5; DB 2; Length 329;

Best Local Similarity 19.6%; Pred. No. 0.068;

Matches 64; Conservative 67; Mismatches 102; Indels 93; Gaps 15;

QY 2 GGGGSIIMKQPKSIINTSQD-FEKRIKIKKV-----INDP 36

Db 6 GYSEKVEKTSKIKRTEEDNLRSREELIKDKKIIKIDSEISRLCLKSTNIFKILNR 65

QY 37 DVKQFLFAHRAELTNMID-EDINVLQAEYKQD--QKHGDKHKAAD-----CPNFVK 85

Db 66 D--EYIKSLKNEITNLMERSSELLVNGVPMVDYLQHNHNSCKDKDGYIGAHKSCGY-KK 122

QY 86 HPELVYNNRIRIRYLOCPCKIKYDEERPEALITSHMQRDTLNAKLKDIYNNRDL 145

Db 123 YIVQLHYDMSDLK-----SLNENNFENENIN-----YYSNRK--- 155

QY 146 DVAMADIDICTAITNGEYV-----KGLYLYGPFGTGSFLLGAIANQLKS 190

Db 156 ---SEDEPTPKKNMEKILISLNTYIKPDSNENMLTFGNSGTGKTFLSHCIADLLE 211

QY 191 K--KYRSTIYLPFIRTLKGGFKDSEFKLHVRREANIIMLDDIGAEEVTPWVRDEV 247

Db 212 RGLVVIYKTSADLAQELKQLQ--FEENPTLEDL--IINDLLIIDLGTQISAFSKTCF 267

QY 248 IGPLHYRMVHELPTFPSSNFDYSEL 273

Db 268 FN-LINKKLISQKMLISSNMLEL 292

Search completed: December 15, 2003, 15:16:45  
Job time: 19.652 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 15:11:40 ; Search time 9.90234 Seconds

(without alignments)  
1486.453 Million cell updates/sec

Title: US-09-689-952-2

Perfect score: 1643

Sequence: 1 MCGGSGIMKQFKSINTSOD.....BRVSLSTPYFLSGENFRNN 313

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt 41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	631	38.4	311	1	DNAI_BACSU
2	137.5	8.4	399	1	DNAI_AQUAE
3	133.5	8.1	313	1	YOMI_BACSU
4	128.5	7.8	457	1	DNAI_BACBP
5	127	7.7	266	1	XKOC_BACSU
6	126	7.7	440	1	DNAI_CAMDE
7	122	7.4	246	1	DNAI_BUCAP
8	117.5	7.2	449	1	DNAI_BACHD
9	117.5	7.2	454	1	DNAI_BUCAL
10	115.5	7.0	246	1	PSMR_BUCAL
11	115	7.0	410	1	PSMR_METTR
12	114	6.9	265	1	ISTB_PSEAE
13	112	6.8	445	1	DNAI_ACHLA
14	111.5	6.8	436	1	PSMR_METKA
15	111	6.8	1208	1	PCPI_SCHPO
16	109.5	6.7	229	1	ISTB_BURCE
17	109	6.6	791	1	LON_CAMDE
18	109	6.6	831	1	LON_HELUP
19	108.5	6.6	454	1	DNAI_BUCAP
20	107.5	6.5	263	1	ISTB_BACFR
21	107.5	6.5	1940	1	MYH3_CHICK
22	107	6.5	451	1	DNAI_PASMT
23	106.5	6.5	251	1	ISTB_BACST
24	105.5	6.4	231	1	ISTB_PSEFL
25	105.5	6.4	432	1	SKOI_SCHPO
26	105	6.4	248	1	YDAV_ECOLI
27	105	6.4	1938	1	MYH6_RAT
28	103	6.3	263	1	YAHM_RHISN
29	103	6.3	450	1	DNAI_MYCCA
30	102.5	6.2	466	1	DNAI_PEWMT
31	102.5	6.2	1939	1	MYH1_HUMAN
32	102	6.2	462	1	DNAI_YERPE
33	102	6.2	464	1	DNAI_SERMA

34	102	6.2	584	1	TRP4_YEAST
35	101	6.1	398	1	PSMR_ARCFU
36	101	6.1	402	1	PSR8_MANGS
37	101	6.1	810	1	YLZ5_CAEEL
38	101	6.1	1938	1	MYH6_MOUSE
39	100.5	6.1	777	1	LON_BUCAL
40	100.5	6.1	794	1	SC18_CANAL
41	100.5	6.1	1938	1	MYSS_CHICK
42	100	6.1	307	1	RYUB_MYCGR
43	100	6.1	403	1	PSR8_DROME
44	100	6.1	643	1	DNAI_STRRE
45	100	6.1	656	1	DNAI_STRCO

## ALIGNMENTS

RESULT 1	ID	DNAI_BACSU	STANDARD	PRT	311 AA.
AC	P06567				
DT	01-JAN-1988	(Rel. 06, Created)			
DT	01-JAN-1988	(Rel. 06, Last sequence update)			
DT	28-FEB-2003	(Rel. 41, Last annotation update)			
DE	Primosomal protein dnaI.				
GN	DNAI.				
OS	Bacillus subtilis.				
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.				
OX	NCBI_TaxId=1423;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=87117549; PubMed=3027671;				
RA	Ogasawara N., Moriya S., Mazza P.G., Yoshikawa H.;				
RT	"Nucleotide sequence and organization of dnaB gene and neighbouring				
RT	genes on the Bacillus subtilis chromosome.";				
RL	Nucleic Acids Res. 14:9989-9999(1986).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=168;				
RX	MEDLINE=97124191; PubMed=8969504;				
RA	Wipat A., Carter N., Brignell C.S., Guy J.B., Piper K.;				
RA	Sanders J., Emerson P.T., Harwood C.R.;				
RT	"The dnaB-pheA (256 degrees-240 degrees) region of the Bacillus				
RT	subtilis chromosome containing genes responsible for stress				
RT	responses, the utilization of plant cell walls and primary				
RL	metabolism.";				
RL	Microbiology 142:3067-3078(1996).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=168;				
RX	MEDLINE=98048467; PubMed=9387221;				
RA	Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;				
RT	"Sequencing and functional annotation of the Bacillus subtilis genes				
RT	in the 200 kb rmb-dnaB region.";				
RL	Microbiology 143:3431-3441(1997).				
RN	[4]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=168;				
RX	MEDLINE=98044033; PubMed=9384377;				
RA	Azeredo V., Bartero M.G., Bessieres P., Bolotin A., Borchert S.;				
RA	Kuntz F., Ogasawara N., Moser I., Albertini A.M., Allioni G.;				
RA	Bouillier S., Brunsch C.V., Caldwell B., Capuano V., Carter N.M.;				
RA	Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.;				
RA	Denzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.;				
RA	Entian K.D., Brington J., Fabret C., Ferrari B., Foulger D.;				
RA	Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.;				
RA	Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.;				
RA	Gutsepri G., Guy B.U., Haga K., Haeck J., Harwood C.R., Hentut A.;				
RA	Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.;				
RA	Joris B., Karamata D., Kasahara Y., Klaetr-Blanchard M., Klein C.;				
RA	Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.;				
RA	Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.;				

RA Lee S.M., Levine A., Liu H., Maeda S., Mauei C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Nockack M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portetelle D., Potwolik S., Prescott A.M.,  
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Rocha B., Roche B., Rose M., Sadale Y.,  
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,  
RA Sekituchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
RA Sorochin A., Taccoti E., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Takakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
RA Tosaou V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Dandhin A.,  
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
RT *subtilis*,"  
RL Nature 390:249-256(1997).  
RN [5]  
RP SEQUENCE OF 1-206 FROM N.A.  
RX MEDLINE=87118226; PubMed=3027697;  
RA Hoshino T., McKenzie T., Schmidt S., Tanaka T., Sueoka N.;  
RT "Nucleotide sequence of *Bacillus subtilis* dnaB: a gene essential for  
RT DNA replication initiation and membrane attachment,"  
RL Proc. Natl. Acad. Sci. U.S.A. 84:653-657(1987).  
RN [6]  
RP SIMILARITY TO DNA.  
RX MEDLINE=92195821; PubMed=1549481;  
RA Koonin E.V.;  
RT "Archaeobacterial virus SSV1 encodes a putative DnaA-like protein,";  
RL Nucleic Acids Res. 20:1143-1143(1992).  
RN [7]  
RP IDENTIFICATION.  
RX MEDLINE=95291463; PubMed=7773414;  
RA Brnard C., Ehrlich S.D.;  
RT "The *Bacillus subtilis* dnaI gene is part of the dnaB operon,";  
RL Microbiology 141:1199-1200(1995).  
CC -1- FUNCTION: PROBABLY INVOLVED IN DNA REPLICATION.  
CC -1- SIMILARITY: SOME, TO DNA FROM VARIOUS BACTERIA.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X04963; CAA28633.1; -;  
DR EMBL; 275208; CAA99605.1; -;  
DR EMBL; AF008320; AAC00359.1; -;  
DR EMBL; 299118; CAB14858.1; -;  
DR EMBL; M15183; AAA22405.1; -;  
DR PIR; B24720; I0BS44.  
DR Subtilin; BG10359; dnaI.  
DR InterPro; IPR003593; AAA\_ATPase.  
DR SMART; SMO0382; AAA; 1.  
KW Primosome; DNA replication; ATP-binding; Complete proteome.  
PT NP\_BIND 168 175  
FT CONFLICT 19 19 K -> N (IN REF. 5).  
FT CONFLICT 24 24 M -> T (IN REF. 5).  
SQ SEQUENCE 311 AA; 3611 MW; A86FC9AB6841264 CRC64;  
Query Match 38.4%; Score 631; DB 1; Length 311;  
Best Local Similarity 43.1%; Pred. No. 4.9e-36;  
Matches 127; Conservative 64; Mismatches 102; Indels 2; Gaps 2;

QY 139 MNHRDLVDVAMADICTAITNGEQYKGLYVFGFTGSKFLGAIYANOLKSKKVRSTII 198  
DB 136 ISDPKSLMFOHVTPDLKSYNETGKKGGLYLYCKFVGKTPMLAAIYANLAEYVSMTIV 195  
QY 199 YLPEFRTLKGKFGKDSFEKFLHRYEANIIMDDIAGAEVTPWVDEVIQPLHYRMYH 258  
DB 196 YLPEFRELKNSLDQDTLEKINMVTPTVLMDDIGAESMTSWVDEVIQVLYQHRMSQ 255  
QY 259 ELPTFPSSNPDVSELEHHLAMTRDGE-EKTKAARIERKSLSTPYFLSGENRRN 312  
DB 256 QLPTFPSSNPDVSELEHHLAMTRDGE-EKTKAARIERKSLSTPYFLSGENRRN 310  
RESULT 2  
ID DNA\_AQUAE STANDARD; PRT; 399 AA.  
AC 066659;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Chromosomal replication initiator protein dnaA.  
GN DNA OR AQ\_322.  
OS Aquifex aeolicus.  
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.  
OX NCBI\_Taxid=63363;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=VF5;  
RC MEDLINE=98196666; PubMed=9537320;  
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujay M., Huber R.,  
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
RT "The complete genome of the hyperthermophilic bacterium *Aquifex*  
RT *aeolicus*,"  
RL Nature 392:353-358(1998).  
RN [2]  
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 76-399.  
RX MEDLINE=22220049; PubMed=12234917;  
RA Erzberger J.P., Pirruccello M.M., Berger J.M.;  
RT "The structure of bacterial DnaA: implications for general mechanisms  
RT underlying DNA replication initiation,"  
RL EMBO J. 21:4763-4773(2002).  
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE INITIATION AND REGULATION  
CC OF CHROMOSOMAL REPLICATION. BINDS TO THE ORIGIN OF REPLICATION;  
CC IT BINDS SPECIFICALLY DOUBLE-STRANDED DNA AT A 9 BP CONSENSUS  
CC (DNA BOX): 5'-TATTC(C/A)A/C(A/A)-3'. DNA BINDS TO ATP AND TO  
CC ACIDIC PHOSPHOLIPIDS (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE dnaA FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AE000683; AAC06612.1; -;  
DR PIR; G70328; G70328.  
DR PDB; 1L8Q; 25-SEP-02.  
DR HAMAP; MF\_00377; -; 1.  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR001957; Bac\_DnaA.  
DR Pfam; PF00308; bac\_dnaA; 1.  
DR PRINTS; PR00051; DNAA.  
DR SMART; SMO0382; AAA; 1.  
DR TIGRFA; TIGR00362; DnaA; 1.  
DR PROSITE; PS01008; DNAA; 1.  
KW DNA replication; DNA-binding; ATP-binding; Complete proteome;  
KW 3D-structure.  
PT NP\_BIND 119 126  
FT CONFLICT 119 126 ATP (POTENTIAL).  
SQ SEQUENCE 399 AA; 46840 MW; C641CB199AA79906 CRC64;

Query Match 8.4%; Score 137.5; DB 1; Length 399;  
 Best Local Similarity 20.2%; Pred. No. 0.009;  
 Matches 48; Conservative 52; Mismatches 69; Indels 69; Gaps 8;

QY 6 SIMKOFKSIINTSODEFKRIEIKKE-----VINDPVKOFLEAHRAELVTAMDEIDN 59  
 DB 5 ALIKIESDYARBELKFE- IKOKGFELFIAPEDREWLET-----IVNTPLEEBR 59  
 QY 60 VIOEKYDQOKHYDGHKFPDCEPFGKCHVELYVNNRIKIRYLOCPCKIKYDEERPEAL 119  
 DB 60 KLEIYVEK-----EKKKVI 75  
 QY 120 IISHMORPTLNK--LKDIYNNHRDLVAMAAADICTAINT-GEQVKGVLVYGFPG 176  
 DB 76 -----KQFLNPKTLENFIVGEGNRL-----AYEVKALELGLSLNPIPIYSGVG 124  
 QY 177 KSFIIIGALINOLKSKKVRSTIYLYLPEFIRTLKGFDSFEKKLHVRANITMLDDI 234  
 DB 125 KTHLQAGNEAKKRGYRIVYSSADPQAQAVHKLKGTINERPNRYKSDLLDDV 182

## RESULT 3

YOAM\_BACSU STANDARD; PRT; 313 AA.  
 ID YOAM\_BACSU  
 AC P45910;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical protein ygam.  
 GN YOAM.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / JH642;  
 RX MEDLINE=95219086; PubMed=7704261;  
 RA "Takemaru K.-I., Mizuno M., Sato T., Takeuchi M., Kobayashi Y.;  
 RT "Complete nucleotide sequence of a skin element excised by DNA  
 RT rearrangement during sporulation in *Bacillus subtilis*.";  
 RL Microbiology 141:323-327(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / JH642;  
 RX MEDLINE=97124195; PubMed=8969508;  
 RA Mizuno M., Maesuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,  
 RA Kobayashi Y.;  
 RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of  
 RT the *Bacillus subtilis* genome containing the skin element and many  
 RT sporulation genes.";  
 RL Microbiology 142:3103-3111(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunze F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Betero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillette S., Brusch C.V., Caldwell B., Capiano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Comercon I.F., Cummings S.J., Daniel R.A.,  
 RA Dentzot F., Devine K.M., Duerenoth A., Ehrlich S.D., Emmerson P.T.,  
 RA Eutlian K.D., Ertington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallazzi A., Galleron N.,  
 RA Ghim S.Y., Glaeser P., Goffeau A., Golligly B.J., Grandi G.,  
 RA Guisepi G., Guy B.J., Haga K., Haele J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koether P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapides A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Maesuda S., Manuel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,

RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Prescan E., Pujic P., Purnelle B., Rapoport G., Ray M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,  
 RA Sato T., Scanlan B., Schleich S., Schroeter R., Scoffone F., Soldo B.,  
 RA Sekiguchi U., Sekowska A., Serr S.U., Serr R., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni A.,  
 RA Toato V., Uchiyama S., Vandenbol M., Vanlier F., Vassarotti A.,  
 RA Viati A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,  
 RA Winers P., Wipac A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein B., Yoshikawa H., Danchin A.,  
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
 RT *subtilis*."  
 RL Nature 390:249-256(1997).  
 RN [4]  
 RP IDENTIFICATION.  
 RX MEDLINE=96084975; PubMed=7489895;  
 RA Medigue C., Moszer I., Viati A., Danchin A.;  
 RT "Analysis of a *Bacillus subtilis* genome fragment using a co-operative  
 RT computer system prototype.";  
 RL Gene 165:GC37-GC51(1995).  
 CC -1- SIMILARITY: TO B.SUBTILIS YQXC AND T.HYDROXYBENTRINE HEMOLYSIN  
 CC TLVA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb.slb.ch](mailto:license@isb.slb.ch)).

DR EMBL; D32216; BAA06927.1; -  
 DR EMBL; D84432; BAA12388.1; -  
 DR EMBL; Z99117; CAB14567.1; -  
 DR PIR; D69945; D69945.  
 DR Subtilast; BG1264; YGAM.  
 DR InterPro; IPR003593; AAA\_Artpase.  
 DR SMART; SM00382; AAA; 1.  
 KW Hypothetical protein: Complete proteome.  
 SQ SEQUENCE 313 AA; 36138 MW; 09208795E310D73A CRC64;

Query Match 8.1%; Score 133.5; DB 1; Length 313;  
 Best Local Similarity 24.4%; Pred. No. 0.013; 90; Indels 83; Gaps 15;  
 Matches 68; Conservative 38; Mismatches 90;

QY 60 VIOEKYDQOKHYDGHKFA-----DCPNFVKGHPVELYVNNRIKIRYLOCPCKIKYD 111  
 DB 16 IMQELREBSARFLGRKQAVSEKAFEDCP-----YCDRGIVVVRVHKDTSWMD 65  
 QY 112 ERFPAELITSHM--QRDTLNKL-----KDIYNN-----HRDLVAMAAADIC 155  
 DB 66 E---QDLMVPPDDMVSBDPLGKVCPTDKASEMDQYKSCQCVRRKRIARLMAASGTT 122  
 QY 156 -----TATNG-----BOVGS-----LYLVGPFNGSGFIIG 182  
 DB 123 EEFKLLFENFTDQKPMIDKAYCAVEYYDFOFKERONSIALQOPSGKTHLT 182  
 QY 183 AIANOL-KSKKVRSTIYLP--EFTRTLKGFDSFEKKLHVRANITMLDDI----- 234  
 DB 183 AINNNLLIKKSH--CMFFPYVSGMDLKAN--DNLKAKLMAKRVEVLFTDLPKPIN 238  
 QY 235 GAEVTPVWRDEVIGPLHYRMVHLELPFGSNFYSEL 273  
 DB 239 GQPRATDW-QVQIOQSVNRYRLNKKPLISSELTIDEI 276

## RESULT 4

DNA\_BUCBP STANDARD; PRT; 457 AA.  
 ID DNA\_BUCBP  
 AC P59567;  
 DT 15-SEP-2003 (Rel. 42, Created)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Chromosomal replication initiator protein dnaX.  
 GN DNA OR BBP012  
 OS Buchnera aphidicola (subsp. Baizongia pistacis).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 CC Enterobacteriaceae; Buchnera.  
 RX NCBI\_TaxID=135842;  
 RP [1]  
 SEQUENCE FROM N.A.  
 RA MEDLINE=2426901; PubMed=12522265;  
 RA Van Ham R.C.H.U., Kamerbeek J., Palacios C., Raueell C., Abascal F.,  
 RA Bastolla U., Fernandez J.M., Jimenez L., Postigo M., Silva F.J.,  
 RA Tamames J., Viguera E., Latorre A., Valencia A., Moran F., Moya A.;  
 RT "Reductive genome evolution in Buchnera aphidicola."  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:581-586(2003).  
 CC -1- FUNCTION: Plays an important role in the initiation and regulation  
 CC of chromosomal replication. Binds to the origin of replication; it  
 CC binds specifically double-stranded DNA at a 9 bp consensus (dnaX  
 CC box): 5'-TATTC(C/A)A(C/A)A-3'. DnaX binds to ATP and to acidic  
 CC phospholipids (By similarity).  
 CC -1- SIMILARITY: Belongs to the dnaX family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; A014016; A026756.1; -  
 DR HAMAP; MF 00377; -; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS01008; DNaA; PALSE NEG.  
 KW DNA replication; DNA-binding; ATP-binding; Complete proteome.  
 FT NP BIND 162 169 ATP (POTENTIAL).  
 SQ SEQUENCE 457 AA; 53365 MW; B2950813FADD4028 CRC64;  
 Query Match 7.8%; Score 128.5; DB 1; Length 457;  
 Best Local Similarity 22.3%; Pred. No. 0.046;  
 Matches 67; Conservative 46; Mismatches 105; Indels 83; Gaps 12;  
 QY 84 KGHVPELYVN-----NRKIRYLQ-----CP-----CKIYDEFEAE 119  
 DB 34 KKNILILVAFNEFSFMKIKNYENLKKLNKCNINTPTLMKICKPKILQKKFNE 93  
 QY 120 ITSHMQORDTLAKDLYNHRDLVMAADICTAINTGEQVKG----- 166  
 DB 94 T-----LKNILINSLT-----YNAVTKISNIYSSSEINTNTYFQNTKQSNQLAFKIYKI 146  
 QY 167 -----LVYGFPGFGKSPILGALINOLKSKKVRSTIYL-----PERIRLKGSPK 213  
 DB 147 AHNPGKNYFPLPLYGSGSGKTHLHAVANITLKIKYIKIYINSENIQMTSLKN 206  
 QY 214 GSEFEKILHREANILMLDDIGAEVTPWRDEV-----IGLLHY-----RMVA 258  
 DB 207 NTIEEPKRYKRVNTLLIDIOFPAYKKSQBELFTINMLLRNQCIIITSDQFQKH 266  
 QY 259 ELPTFFSSNDYSELHHLAMTDBGEKRYAARI---ERVKSLSTPY---FLSGENFRN 312  
 DB 267 GIETRLKSRF-----ECGLTIRIDPDLNRTKILKKSHIYDINLSKYVAFIANKLKS 321  
 QY 313 N 313  
 DB 322 N 322

DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Phase-1like element PBX protein xkdc.  
 GN xkdc  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 RX NCBI\_TaxID=1423;  
 RP [1]  
 SEQUENCE FROM N.A.  
 RA STRAIN=168 / S0113;  
 RA MEDLINE=94364963; PubMed=8083174;  
 RA McDowell G.E., Wood H., Devine K.M., McConnell D.U.;  
 RT "Genetic control of bacterial suicide: regulation of the induction of  
 RT PBX in Bacillus subtilis."  
 RL J. Bacteriol. 176:5820-5830(1994).  
 CC [2]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=168;  
 RA Krog S., O'Reilly M., Nolan N., Devine K.M.;  
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Berteiro M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Boriss R., Bourstier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Bruchet C.V., Caldwell B., Capiano V., Carter N.M.,  
 RA Choi S.K., Codani J.U., Conneron I.F., Cummings N.J., Daniel R.A.,  
 RA Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Eutian K.D., Errington J., Fabret C., Ferrati E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghim S.Y., Glaeser P., Goffeau A., Golightly E.J., Grandt G.,  
 RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Bianchard M., Klein C.,  
 RA Kobayashi Y., Koester P., Koningsstein G., Krog S., Kumano M.,  
 RA Kurita K., Lapdue A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogawa A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetlelle D., Porwollik S., Prescott A.M.,  
 RA Prescan E., Pujic P., Purnelle B., Rapoport G., Ray M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,  
 RA Sato T., Scanlan B., Schleich S., Schroeder R., Scoffone F.,  
 RA Sekiguchi J., Sekwaka A., Serró S.J., Serró P., Shin B.S., Soldo B.,  
 RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni K.,  
 RA Tosato V., Uchiyama S., Vandembol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weltzenegger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zanchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
 RT subtilis."  
 RL Nature 390:249-256(1997).  
 CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTIONAL ANTITERMINATOR.  
 CC -1- SIMILARITY: STRONG, TO B. SUBTILIS Y0AM.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z34287; CA84044.1; -  
 DR EMBL; Z70177; CA84054.1; -  
 DR EMBL; Z99110; CA83110.1; -  
 DR PIR; I40411; I40411.  
 DR Subtilist; BG10996; xkdc.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR SMART; SM00382; AAA; 1.  
 KW ATP-binding; Complete proteome.

FT NP\_BIND 124 131 ATP (POTENTIAL).  
 FT CONFLICT 239 K -> R (IN REF. 1).  
 SQ SEQUENCE 266 AA; 30534 MW; 9929C91E9D655AA CRC64;  
 Query Match 7.7%; Score 127; DB 1; Length 266;  
 Best Local Similarity 31.0%; Pred. No. 0.031;  
 Matches 39; Conservative 25; Mismatches 44; Indels 18; Gaps 7;  
 QY 155 CNA--ITNEQVYK-----GLYXPGRTGKSPFLGAIANQKSKYRSTIYLP--EFTIR 205  
 DB 101 CKEVYADVEQIDCKRNSIALGPGSGKTHLTAAHEL-WRTCYVPVITYPVEGFT 159  
 QY 206 TLKGGKDGSPFKLHREBEANILMDI-----GAEVTPVWVDEVIGELIYRWHEL 260  
 DB 160 DLKND--ALTEKANKRMQADVLPIDDLKPYNGKPRATDQLEQMY-SVNTRYLANK 216  
 QY 261 PTFESS 266  
 DB 217 PILSS 222  
 RESULT 6  
 DNAA CAMJE STANDARD; PRT; 440 AA.  
 ID DNAA CAMJE  
 AC 09JB0;  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Chromosomal replication initiator protein dnaA.  
 GN DNAA OR C30001.  
 OS Campylobacter jejuni.  
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;  
 OC Campylobacteraceae; Campylobacter.  
 OX NCBI\_Taxid=197;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCTC 11168;  
 RX MEDLINE=20150912; PubMed=10688204;  
 RA Parkhill J., Wen B.W., Mungall K., Kelsey J.M., Churcher C.,  
 Basham D., Chillingworth T., Davies R.M., Feltham T., Holtroyd S.,  
 Jagsis K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,  
 Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,  
 Whitehead S., Barrett B.G.;  
 RA "The genome sequence of the food-borne pathogen Campylobacter jejuni  
 reveals hypervariable sequences.";  
 RL Nature 403:665-668(2000).  
 RT -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE INITIATION AND REGULATION  
 OF CHROMOSOMAL REPLICATION. BINDS TO THE ORIGIN OF REPLICATION;  
 IT BINDS SPECIFICALLY DOUBLE-STRANDED DNA AT A 9 BP CONSENSUS  
 (DNAA BOX): 5'-TATC(C/A)A(C/A)A-3'. DNAA BINDS TO ATP AND TO  
 ACIDIC PHOSPHOLIPIDS (BY SIMILARITY).  
 CC -1- SIMILARITY: Belongs to the dnaA family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL, AL139074, CAB72494.1, -  
 DR PIR, D81415, D81415.  
 DR HAMAP, MF\_00377, -; 1.  
 DR InterPro, IPR003593; AAA\_ATPase.  
 DR InterPro, IPR001957; Bac\_DnaA.  
 DR Pfam, PF00308; bac\_dnaA\_1.  
 DR PRINTS, PRO0051; DNAA.  
 DR SMART, SM00382; AAA; 1.  
 DR TIGRFAMs, TIGR00362; DnaA; 1.  
 DR PROSITE, PS01008; DNAA; FALSE NEG.  
 DR DNA replication; DNA-binding; ATP-binding; Complete proteome.  
 FT NP\_BIND 143 150 ATP (POTENTIAL).

SQ SEQUENCE 440 AA; 49686 MW; A3FAB90CCE0DEBD7 CRC64;  
 Query Match 7.7%; Score 126; DB 1; Length 440;  
 Best Local Similarity 21.3%; Pred. No. 0.067;  
 Matches 64; Conservative 51; Mismatches 93; Indels 92; Gaps 13;  
 QY 40 QPLEHRAELTAMIDEDLVLOEYKQKHVDGKHFADCPFKGVKGVPELYDN--NRI 97  
 DB 5 QILENKLKELSENREYENTISNKLKFNKQSK-----ADLVFNAPNEL 46  
 QY 98 KIRYLQCPCKIRYDEERFAELITSHNQDTLNAKLDI-----YNNHRLDVA--- 148  
 DB 47 MAKFIQ---TKGKK-----IAHFYVGSGNAIINIQAQAKQSKSTKIDIAHK 95  
 QY 149 -----MAADDICATINSGOVKLY-----LYGPGTGSITLA 183  
 DB 96 AGSTILNPSFTFESFVVDGNSNRYAGACAIHAKDGLVNPIFYGPTGKTHLQA 155  
 QY 184 IANOLSKKVRSTIYL--PEFIRTLKGGFKDGSFEKKLHREBEANILMDI GAEVTP 241  
 DB 156 VGN--ASLEMKGKVIYATSENINDFTSLKNGSLDKHEKTRNCVDVLLIDVQPIGKTD 213  
 QY 242 WVRDEVIGPLHYRWVHELPTFESSNPDYSLEHH--LAWTRDGEKTKAARIIVERVS 298  
 DB 214 KIQEE-----FF--FIFNEIKNDGQIIMTSNDPNNLKG-ITERLKS 253  
 RESULT 7  
 DNAC BUCAP STANDARD; PRT; 246 AA.  
 ID DNAC BUCAP  
 AC 08KA79;  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE DNA replication protein dnaC.  
 GN DNAC OR B08G022.  
 GN Buchnera aphidicola (subsp. Schizaphis graminum).  
 OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Buchnera.  
 OX NCBI\_Taxid=98794;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=22084549; PubMed=12089438;  
 RX Tamás I., Klaesson L., Canbaeck B., Naeslund A.K., Eriksson A.-S.,  
 Wernegreen J.V., Sandstrom J.P., Moran N.A., Andersson S.G.E.;  
 RT "50 million years of genomic stasis in endosymbiotic bacteria.";  
 RL Science 296:2376-2379(2002).  
 CC -1- FUNCTION: THIS PROTEIN IS REQUIRED FOR CHROMOSOMAL REPLICATION. IT  
 FORMS, IN CONCERT WITH DNA PROTEIN AND OTHER PREPROMING PROTEINS  
 DNAT, N', N'', A PREPROMING PROTEIN COMPLEX ON THE SPECIFIC SITE  
 OF THE TEMPLATE DNA RECOGNIZED BY PROTEIN N' (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE DNAC FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL, AE014077; AAM67594.1, -  
 DR DNA replication; Primosome; Complete proteome.  
 FT SITE 69  
 FT PROBABLY INVOLVED IN THE INTERACTION  
 WITH THE DNAB PROTEIN (BY SIMILARITY).  
 SQ SEQUENCE 246 AA; 28497 MW; BF7E2A9BC2ADBBD CRC64;  
 Query Match 7.4%; Score 122; DB 1; Length 246;  
 Best Local Similarity 23.6%; Pred. No. 0.064;  
 Matches 48; Conservative 39; Mismatches 86; Indels 30; Gaps 7;  
 QY 111 DERREAEALITSH--MORDTLNAKLKDIYMN-----HRDRLDVNAADDICTA 157

DB 35 ECGRLSESSILRENMKQVILGRSGIRRELYMNCSPENYKIEHGQKYLKAKRYAEE 94  
 QY 158 ITNGEOVKGLYVPGFGKSPFLIGAIANOLKSKVSTIIVPEFIRTLKGGFKDOS-- 215  
 DB 95 P--NENIASIFSGRPGTGNKSLASATGNVILHGSILVTVADLSNNKKGFTSGTSNT 152  
 QY 216 -FEKQHRVBEANIMLDDIGAEVTPWVDEVIPLHRYMHELPTFPSSNEDYSELE 274  
 DB 153 TEENLHNLSVLDLMDIEIGMQTESRYEK-VIINQVDRSSSKSTGMLSNLDHRGMK 211  
 QY 275 HHLAMTRDGEKTKAARIIRVK 297  
 DB 212 NLL-----GR-----RVIDRR 223  
 RESULT 8  
 ID DNA\_BACHD STANDARD; PRT; 449 AA.  
 AC Q9RC2;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Chromosomal replication initiator protein dnaa.  
 GN DNAA OR BH0001.  
 OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=86665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RX MEDLINE=99356711; PubMed=10427704;  
 RA Takami H., Masui N., Nakasone K., Horikoshi K.;  
 RA Fujii F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
 RA Horikoshi K.;  
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
 halodurans and genomic sequence comparison with Bacillus subtilis.";  
 RL Nucleic Acids Res. 28:4317-4331(2000).  
 CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE INITIATION AND REGULATION  
 OF CHROMOSOMAL REPLICATION. BINDS TO THE ORIGIN OF REPLICATION;  
 CC IT BINDS SPECIFICALLY DOUBLE-STRANDED DNA AT A 9 BP CONSENSUS  
 CC (DNA BOX): 5'-TTATC(C/A)A(C/A)-3'. DNAA BINDS TO ATP AND TO  
 CC ACIDIC PHOSPHOLIPIDS (BY SIMILARITY).  
 CC -1- SIMILARITY: Belongs to the dnaa family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: AB013492; BAA82685.1; -  
 CC EMBL: AP001507; BAB03720.1; -  
 CC PIR: A83650; A83650.  
 CC HAMAP: MF\_003777; - 1.  
 CC InterPro: IPR003593; AAA\_ATPase.  
 CC InterPro: IPR001957; Bac\_Dnaa.  
 CC Pfam: PF00308; bac\_dnaa; 1.  
 CC PRINTS: PR00051; DNAA.  
 CC SMART: SM00382; AAA; 1.  
 CC TIGRfam: TIGR00362; Dnaa; 1.  
 CC PROSITE: PS01008; DNAA; 1.  
 CC KEGG: K00001; DNAA; 1.  
 KW DNA replication; DNA-binding; ATP-binding; Complete proteome.  
 FT NP\_BIND 153 160

SQ SEQUENCE 449 AA; 51263 MW; D75C00BBF110A614 CRC64;  
 Query Match 7.2%; Score 117.5; DB 1; Length 449;  
 Best Local Similarity 20.8%; Pred. No. 0.28;  
 Matches 67; Conservative 49; Mismatches 107; Indels 99; Gaps 16;  
 QY 21 FEKRIKIKKEVINDEPV-----KQFLAHRALTNAMIDEDLNLQAEYKDOOKHY 71  
 DB 24 FETMLQNTANSIEDSTIITTAPEFARFVLEHGYDELLSETID-DLTGVRLV----- 75  
 QY 72 DGHKPADCPNFKGVHVELVDNNRIRKRYLQCPCKIKTDEEFKALITSHMORDTLN 121  
 DB 76 -----PKFV-----IPTSLD-----EPFVQELKKPKQPPAQN 105  
 QY 132 AKKDIYNNHRRDLVDVMA-----ADDICATITNGEOVKGLYVPGFGKSPFLIGA 183  
 DB 106 GEPNMMMDKTYFDYFVIGSGNRFHAAASLVAEPAPAKAYNPLFYGGVGLKTHLMA 165  
 QY 184 IANOLSKKVRSTIIVP-----EFIRTLKGGFGKSPFKLHVRANIMLDDI--- 234  
 DB 166 IGHVMDHNPNAKVYVLSSEKFTNEFINARON-KAVNFRNK---YRNVDLLIDDIQFL 221  
 QY 235 -GAEEVTPWVDEVIQPL-LHY-----RMVHELPTF--FSSNFDIS----- 271  
 DB 222 AGKEQ-----TOEEFFFTFNALEDNKKQIVISSDRPPEKIPTEDRLSRFEWGLITDITP 277  
 QY 272 -ELHHLAMTRDGEKTKAARI 292  
 DB 278 PDLERIRALIR--KKAQENL 296  
 RESULT 9  
 ID DNA\_BUCAI STANDARD; PRT; 454 AA.  
 AC P57128;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Chromosomal replication initiator protein dnaa.  
 GN DNAA OR BU012  
 OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum  
 OS symbiotic bacterium).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Buchnera.  
 OX NCBI\_TaxID=118099;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Tokyo 1998;  
 RX MEDLINE=20445173; PubMed=10993077;  
 RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;  
 RA "Genome sequence of the endocellular bacterial symbiont of aphids  
 RT Buchnera sp. APS.";  
 RL Nature 407:81-86(2000).  
 CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE INITIATION AND REGULATION  
 CC OF CHROMOSOMAL REPLICATION. BINDS TO THE ORIGIN OF REPLICATION;  
 CC IT BINDS SPECIFICALLY DOUBLE-STRANDED DNA AT A 9 BP CONSENSUS  
 CC (DNA BOX): 5'-TTATC(C/A)A(C/A)-3'. DNAA BINDS TO ATP AND TO  
 CC ACIDIC PHOSPHOLIPIDS.  
 CC -1- SIMILARITY: Belongs to the dnaa family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: AB001119; BAB12740.1; -  
 CC HAMAP: MF\_003777; - 1.  
 CC InterPro: IPR003593; AAA\_ATPase.  
 CC InterPro: IPR001957; Bac\_Dnaa.  
 CC Pfam: PF00308; bac\_dnaa; 1.

DR PRINTS; PRO0051; DNAA.  
 DR SMART; SM00382; AAA; 1.  
 DR TIGR00362; DNAA; 1.  
 DR PROSITE; PS01008; DNAA; 1.  
 DR DNA replication; DNA-binding; ATP-binding; Complete proteome.  
 DR NP BIND 159 166 ATP (POTENTIAL).  
 DR SEQUENCE 454 AA; 52931 MW; 1F5299FAC2751213 CRC64;

Query Match 7.2%; Score 117.5; DB 1; Length 454;  
 Best Local Similarity 22.1%; Pred. No. 0.28;  
 Matches 53; Conservative 46; Mismatches 86; Indels 55; Gaps 10;

QY 25 IEKIKKEVINDPDKVQFLBAHRAELTNAMIDEDLVYQEKDQKKYDGHKFAQCPN-FV 83  
 DB 10 LRLQDELN-TEFSWIRSLKAKLN-----NILETY-----ANKKV 47  
 QY 84 KGHVPELVYDNNRIKIRYLQCPG-----KIKYDEEFPEALITSHMQBDTLNAKIKDIY 138  
 DB 48 LEWVKDKYL-NHLK-KILODYCGTNSPLIKFEIYQIYKEMKLNINNNNNKNEKLIM 104  
 QY 139 MN-----HNRDLVMAAADICTAINT-----GEQVGLYIYGFPGG 176  
 DB 105 SNIPKFNLSYSNNIKRYNPNFVFGSKNQLARSAFOAARNPGNSYNPLFLYGTGIG 164  
 QY 177 KSFILGAINOLSKSKVRSSTIYLPE--FIRTLKGGFKDGSFEKKLHVRVREANITLMDI 234  
 DB 165 KTHLMAIGNEILSYKIDIKIIPNSECFQDMVKALKNNATKFKLYRSVDALIIDI 224

## RESULT 10

DNAC BUCAI STANDARD; PRT; 246 AA.

AC P57134;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE DNA replication protein dnac.  
 GN DNAC OR BU021.  
 OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum  
 OS symbiotic bacterium).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Buchnera.  
 NCBI\_Taxid=118099;

RP SEQUENCE FROM N.A.  
 RC STRAIN=Tokyo 1998;  
 RX MEDLINE=20445173; PubMed=10993077;  
 RA Shigenobu S., Matanabe H., Hattori M., Sakaki Y., Ishikawa H.;  
 RT "Genome sequence of the endocellular bacterial symbiont of aphids  
 RT Buchnera sp. APS";

RL Nature 407:81-86(2000).  
 CC -1- FUNCTION: THIS PROTEIN IS REQUIRED FOR CHROMOSOMAL REPLICATION. IT  
 CC FORMS, IN CONCERT WITH DNAB PROTEIN AND OTHER PREPRIMING PROTEINS  
 CC DNAT, N, N', A PREPRIMING PROTEIN COMPLEX ON THE SPECIFIC SITE  
 CC OF THE TEMPLATE DNA RECOGNIZED BY PROTEIN N' (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE DNAC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF001118; BAB12748.1;  
 CC DNA replication; Primosome; Complete proteome.  
 CC SITE 69 69 WITH THE DNAB PROTEIN (BY SIMILARITY).  
 CC SEQUENCE 246 AA; 28444 MW; C2E56A3BCD1A68F CRC64;

Query Match 7.0%; Score 115.5; DB 1; Length 246;  
 Best Local Similarity 22.5%; Pred. No. 0.19;

Matches 49; Conservative 40; Mismatches 88; Indels 41; Gaps 8;

QY 107 KIKYDEE-----RFEALITSHH-----MORDTLNAKLIYNN-----HR 142  
 DB 20 KRFPNDEDLAMNOEGGLSSSESTIREKNAKMRVLRGSSIRRELYNMSFDPNYKIEHD 79  
 QY 143 DRLDVAAMAADICTAINTGEOYKGLYIYGFPGTGSFIIIGALANOLSKKRVSTIYLPE 202  
 DB 80 GGRKVLKASKRYAEF--NENIASFISGKPGTGNHLSAIGNYILHKGSIILVTVAD 137  
 QY 203 FIRTLKGGFKDGS---FEKKLHVRVREANITLMDIGAEVTVWVDEYIGPLIHYRVH 259  
 DB 138 LMSNNKGFSGSGNSTEETLHDLSSVDLMDIGMGTESRYEK-VIIHQIVDRSSSK 196  
 QY 260 LPTFPSSNDYSELHNLMTDGEKTKAARIIIRVK 297  
 DB 197 RSTGMLSNDRKMSL-----GE-----RVIDMR 223

## RESULT 11

PSMR METH STANDARD; PRT; 410 AA.

AC Q26824;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Proteasome-activating nucleotidase (Proteasome regulatory subunit).  
 GN PAN OR MTH728.  
 OS Methanobacterium thermoautotrophicum.  
 OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;  
 OC Methanobacteriaceae; Methanothermobacter.  
 NCBI\_Taxid=187420;

RP SEQUENCE FROM N.A.  
 RC STRAIN=Delta H;  
 RX MEDLINE=98037514; PubMed=9371463;  
 RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,  
 RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,  
 RA Harrison D., Hoang L., Keagle P., Lumm W., Rother B., Qiu D.,  
 RA Spadafora R., Vicare R., Wang Y., Wleczkowski J., Gibson R.,  
 RA Jiwani N., Caruso A., Bush D., Sater H., Patwell D., Prabhakar S.,  
 RA McDougall S., Shiner G., Goyal A., Pietrowski S., Church G.M.,  
 RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;  
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
 RT deltaH: functional analysis and comparative genomics";  
 RL J. Bacteriol. 179:7135-7155(1997).  
 CC -1- FUNCTION: Required for the ATP- or GTP-dependent degradation of  
 CC proteins, but not small peptides, by the 20S proteasome (By  
 CC similarity).

CC -1- SUBUNIT: Homohexamer (Potential).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF000852; AAB85233.1;  
 CC PIR; G6197; G69187.  
 CC HAMAP; MF\_00553; -; 1.  
 CC InterPro; IPR005937; 268\_P45.  
 CC InterPro; IPR003593; AAA\_Arase.  
 CC InterPro; IPR003959; AAA\_Arase.  
 CC InterPro; IPR003959; AAA\_Arase.  
 CC Pfam; PF00004; AAA; 1.  
 CC SMART; SM00382; AAA; 1.  
 CC TIGR00362; DNAA; 1.  
 CC PROSITE; PS00674; AAA; 1.  
 CC Proteasome; ATP-binding; Complete proteome.

FT NP BIND 192 199 ATP (POTENTIAL).  
 SQ SEQUENCE 410 AA; 4611 MM; 38261703F836AD4 CRC64;  
 Query Match 7.0%; Score 115; DB 1; Length 410;  
 Best Local Similarity 22.3%; Pred. No. 0.37;  
 Matches 59; Conservative 50; Mismatches 101; Indels 54; Gaps 12;

16 NTSDQFKEKIEKIEV--INDPDKO---FLEHRAELTNAMIDDLNVL--QEVYDOQ 68  
 3 NNSQNVAKIEDLCKEYRLMKESKTKRNLMLKRLKLEKDLIERRLDRVYSKR 62  
 69 KHYDGHKADCPNFKGVPELVYNNRIKIRYQCP--CKIKYDEEFELITSHMOR 127  
 63 GEIERFR---TPPLVIAVTEV--LDDHRAVAKSTTGPHFVINYSRFIDRKQLEGRVAL 118  
 128 DTLNAKIDLYMNRDLVDVMAAD---DICTAITNG--EQV----- 164  
 119 NQOTFSIVDVPSEKDVVTGMEVEKPDVSEQIGLEQVREKVTLPKLPFLFE 178  
 165 -----KGLVYGPFGTSGKSPILGAIANOLKS---KKVSTII--YLPEFIRTLKGSPK 212  
 179 KIGIEPPKGVLYVPGTGTILAKAVAHETNAFPIKIVASEFVRKXIGEGARLVGVF-- 237  
 213 DGSFEKLRHVRREANIIMDDIGA 236  
 238 -----ELAKEKSPSIIIFIDEIDA 255

RESULT 12  
 ISTD\_PSEAE STANDARD; PRT; 265 AA.  
 AC P15026;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Insertion sequence IS21 putative ATP-binding protein.  
 GN ISTD.  
 OS Pseudomonas aeruginosa.  
 OC Plasmid R68.45.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=69218951; PubMed=2540414;  
 RA Reimann C., Moore R., Little S., Savioz A., Willetts N.S., Haas D.;  
 RT "Genetic structure, function and regulation of the transposable  
 element IS21.";  
 RL Mol. Gen. Genet. 215:416-424(1989).  
 RN [2]  
 RP REVISION TO 283.  
 RA Berger B.;  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: THE ISTD PROTEIN, ONE OF TWO PROTEINS EXPRESSED ONLY  
 CC WHEN THERE IS A TANDEM REPEAT OF THE IS21 INSERTION SEQUENCE, IS  
 CC NECESSARY FOR THE TRANSDUCTION OF PLASMIDS WITH THAT TANDEM  
 CC REPEAT.  
 CC -1- SIMILARITY: BELONGS TO THE IS21/IS1162 PUTATIVE ATP-BINDING  
 CC PROTEIN FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X14793; CA33899.2; -;  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR InterPro: IPR002611; ISTD\_ATPbind.  
 DR Pfam: PF01695; ISTD.1.  
 DR SMART: SM00382; AAA.1.

KW Plasmid; Transposable element; ATP-binding.  
 FT NP BIND 104 111 ATP (POTENTIAL).  
 SQ SEQUENCE 265 AA; 30528 MM; 9047ADC5E3A12A9 CRC64;  
 Query Match 6.9%; Score 114; DB 1; Length 265;  
 Best Local Similarity 26.1%; Pred. No. 0.26;  
 Matches 42; Conservative 36; Mismatches 53; Indels 30; Gaps 8;

162 EOVKGLVYGPFGTSGKSPILGAIANOLKSKKRVSTIIVP--EFIRTLKGFGDGSPEKK 219  
 95 ESENENVILGPPGVGKTHL--AIALGVKAVDAGRVLVFLDLRLITLMKAKQENRLERQ 152  
 220 LHRVREANILIMDDIGAEEVTPWVRDEVIGPLAHYVNH---ELPTFESSN--PDYS 272  
 153 LQQLSYARVLIIDEIG---YLPWNRDEA---SLFFRLNRRYKASIIILSNKGADWGE 206  
 273 L--EHLNLTROGEKTKARIIRKYSLSLTPFLSGENFR 311  
 207 MFGDHYL-----ATAILDRLLHSHSTLNLNGESYR 236

RESULT 13  
 DNAA\_ACHLA STANDARD; PRT; 445 AA.  
 ID DNAA\_ACHLA  
 AC Q9KDH8;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Chromosomal replication initiator protein dnaA.  
 GN DNAA.  
 OS Acholeplasma laidlawii.  
 OC Bacteria; Firmicutes; Mollicutes; Acholeplasmatales;  
 OC Acholeplasmataceae; Acholeplasma.  
 OX NCBI\_TaxID=2148;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Pg-8B;  
 RX MEDLINE=20242229; PubMed=10779957;  
 RA Taganov K.D., Gushchin A.E., Akopian T.A., Oparina N.Y.,  
 RA Abramychcheva N.Y., Govorov V.M.;  
 RT "Analysis of genes, coding for DNA gyrase from the mycoplasma  
 RT Acholeplasma laidlawii Pg-8B.";  
 RL Mol. Biol. (Mosk) 34:292-299(2000).  
 CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE INITIATION AND REGULATION  
 CC OF CHROMOSOMAL REPLICATION. BINDS TO THE ORIGIN OF REPLICATION;  
 CC IT BINDS SPECIFICALLY DOUBLE-STRANDED DNA AT A 9 BP CONSENSUS  
 CC (DNAA BOX): 5'-TTATC(C/A)A(C/A)A-3'. DNAA BINDS TO ATP AND TO  
 CC ACIDIC PHOSPHOLIPIDS (BY SIMILARITY).  
 CC -1- SIMILARITY: Belongs to the dnaA family.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF248639; AAF75986.1; -;  
 DR HAMAP: MF\_00377; -; 1.  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR InterPro: IPR001957; Bac\_DnaA.  
 DR Pfam: PF00308; bac\_dnaA.1.  
 DR PRINTS: PR00051; DNAA.  
 DR SMART: SM00382; AAA.1.  
 DR TIGRFAMs: TIGR00362; DnaA.1.  
 DR PROSITE: PS01008; DNAA.1.  
 KM DNA replication; DNA-binding; ATP-binding.  
 FT NP BIND 144 151  
 SQ SEQUENCE 445 AA; 51046 MM; A6045EB53F8C9DB3 CRC64;  
 Query Match 6.8%; Score 112; DB 1; Length 445;  
 Best Local Similarity 22.6%; Pred. No. 0.67;

Matches 44; Conservative 36; Mismatches 85; Indels 30; Gaps 9;

QY 58 LNVLYQYKQOQKHYYDGHKFPADCPN-PYKGVHPELYVDN-NRIRIYLOQCPCKIKY-DEER 114  
 Db 29 LPVTSFVKOD---NGLITVWVANEFLKRNINLYAKINELTKYSSTVRLKFPVSOER 84  
 QY 115 FEALITSHMORDTLNKLKDIY-----MNRDRDLVMAAADICTAINTGEQVKG 166  
 Db 85 VIEEPVADRRLTIDYQGNLNTSTYTPDSFVVGKSNMFAFMAKVAHDHPAVAN-----P 139  
 QY 167 LYIYGPFGTGSFTILGAINOLSKKRYSTIYIY-----LPEFIRTL-KGPFQDSEFKK 219  
 Db 140 FYIFGVGVLGKTHMQLAGNYIILNDVYKRLVYKADNFIEDVSLSRKNTKEEFNAK 199  
 QY 220 LHRVREANTILMLDDI 234  
 Db 200 ---YKQIDVILVDDI 211

RESULT 14  
 PSMR\_METKA STANDARD; PRT; 436 AA.

AC Q8TX03;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Proteasome-activating nucleotidase (Proteasome regulatory subunit).  
 GN PAN OR MK0878.  
 OS Methanopyrus kandleri.  
 OC Archaea; Euryarchaeota; Methanopyrii; Methanopyrales; Methanopyraceae;  
 OC Methanopyrus.  
 OX NCBI\_TaxId=2320;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AV19 / DSM 6324 / JCM 9639;  
 RX MEDLINE=21927647; PubMed=11930014;  
 RA Slesarev A.I., Mezheva K.V., Makarova K.S., Polushin N.N.,  
 RA Shchepochina O.V., Shakhova V.V., Belova G.I., Aravind L.,  
 RA Natale D.A., Rogozin I.B., Tatucov R.L., Wolf Y.I., Stetter K.O.,  
 RA Malykh A.G., Koonin E.V., Kozlyavkin S.A.;  
 RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19  
 and monophyly of archaeal methanogens.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).  
 CC -1- FUNCTION: Required for the ATP- or CTP-dependent degradation of  
 CC proteins, but not small peptides, by the 20S proteasome (by  
 CC similarity).  
 CC -1- SUBUNIT: Homohexamers (Potential).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AE010377; AAM02091.1; -  
 CC HAMAP: MF\_00553; -1.  
 CC InterPro: IPR005937; 26S\_P45.  
 CC InterPro: IPR003593; AAA\_ATPase.  
 CC InterPro: IPR003958; AAA\_ATPase\_cent.  
 CC InterPro: IPR003960; AAA\_sub.  
 CC Pfam: PF00004; AAA; 1.  
 CC SMART: SM00382; AAA; 1.  
 CC TIGRPFAM: TIGR01242; 26Sp45; 1.  
 CC PROSITE: PS00674; AAA; 1.  
 CC Proteasome; ATP-binding; Complete proteome.  
 KM NP\_BIND 220 227 ATP (POTENTIAL).  
 FT SEQUENCE 436 AA; 49767 MW; AE150086616A02ED CRC64;  
 Query Match 6.84; Score 111.5; DB 1; Length 436;

Best Local Similarity 21.5%; Pred. No. 0.71;  
 Matches 59; Conservative 55; Mismatches 100; Indels 61; Gaps 13;

QY 5 QSIMKO-FKSIINTSODFEKRIIEKIKGYINDPVKQFIEAR--AELTNMIDDLAV 60  
 Db 27 BELKKEFFRLBELKELKRAHEKRIEARRRRKLEKELMEDEKALEELRRREVM 86  
 QY 61 LQEVYKQOQKHYYDGHKFPADCPNFPYKGVHPELYVDNNRIRIYLOQCP---CKIKYDEEFP 117  
 Db 87 EKIRSLQF-----MKRPLIVGTVEI-LDDGVYKASSTGPFVSNVSTVDREL 138  
 QY 118 ELITSHMORDTLNKLKDIYNNHRDRDLVMAAAD---DICTAINTG---EQV----- 164  
 Db 139 EPGANVALNQSW--AVVDVLPSEKDSRYLAMEVDSPDVSYDDIGLEQREIREV 196  
 QY 165 -----KELYGPFGTGSFTILGAINOLSKKRYSTIYIYPEFIRTL 207  
 Db 197 KPLKEPELEKYGVEEPKGVLLYGPPTGKTLLAKVANHAADATFIR---LAAPELVOKF 253  
 QY 208 KGFQDGS-----PEKLRVREANTILMLDDIGA 236  
 Db 254 IG---EGARLVRELF--LAREKAPSIITFIDEIDA 283

RESULT 15  
 PCP1\_SCHPO STANDARD; PRT; 1208 AA.

AC Q92351;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Spindle pole body protein pcpl1.  
 GN PCP1 OR SPAC69.06C.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OX NCBI\_TaxId=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.  
 RC MEDLINE=21852775; PubMed=11864908;  
 RA Flory M.R., Morphey M., Joseph J.D., Means A.R., Davis T.N.;  
 RT "Pcp1p, a Spo11p-related calmodulin target at the centrosome of the  
 RL fission yeast Schizosaccharomyces pombe.";  
 RL Cell Growth Differ. 13:47-58(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouras J., Peat N., Hayles J., Baker S., Baaham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Felkewell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Nibbelc D., Ocell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Voiclaert G., Aert R., Robben J., Grymouprez B.,  
 RA Wellens I., Vanstreels B., Rieger M., Schaefer M., Muller-Auer S.,  
 RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,  
 RA Beyer P., Zimmermann W., Wedler H., Wambut R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucac M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potaschkin J.,  
 RA Shpakovski G.V., Usery D., Barrett B.G., Nurse P.;

```

RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -1- FUNCTION: Spindle pole body component that binds calmodulin.
CC Overexpression of pcpl causes the formation of supernumerary SPB-
CC like structures and disrupts both mitotic spindle assembly and
CC chromosome segregation.
CC -1- SUBCELLULAR LOCATION: Spindle pole body.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, Z81317, CAB03608.1, -.
DR EMBL, AF348506, AAK31344.1, -.
DR PIR, T39068, T39068.
DR GeneDB Spombe; SPAC6G9.06c; -.
KM Coiled coil.
FT DOMAIN 151 375 COILED COIL (POTENTIAL).
FT DOMAIN 387 803 COILED COIL (POTENTIAL).
FT DOMAIN 874 1091 COILED COIL (POTENTIAL).
FT DOMAIN 1177 1204 COILED COIL (POTENTIAL).
SQ SEQUENCE 1208 AA; 140763 MW; 70264159ADD42424 CRC64;

Query Match 6.8%; Score 111; DB 1; Length 1208;
Best Local Similarity 20.6%; Pred. No. 2.6;
Matches 72; Conservative 52; Mismatches 113; Indels 112; Gaps 16;

QY 4 GOSIMKOPKSIINTSQDFERIEKIKKEVINDPVKQFLBAHRAELTNAMIDE----- 56
DB 693 GQKLASAFELMSNKKQALFKYSSLKNELINAQNLDRREBELSELKLFEEKIRSGS 752
QY 57 -----DIANL-QEYKD--QOKHYDGKFA--DCPNFVKGHVPELYVDNNRIKIRY- 101
DB 753 NNDIEKKKEINVNLSELDKLAQIRHLESDKMELDKLVHHILNRGIEBANIEENAVKRLC 812
QY 102 LQPCP-----KIKYDEERF-----EALITSHHMQR 127
DB 813 LLMGCDYSSVSIQIYVSQIEHFVNOQIQTRSLKQELRHDFVQPSGKKEQL--SRSFEK 870
QY 128 DTINAKLKDIYNNHRDLDAVMAADICTA---ITNGEQVKGILYGFPGTGSFTIGAI 184
DB 871 FGIETETKHILAQRRN-NVSEKNDLENAAQKFPSSPDRKNGLYLPSHTSK----- 922
QY 185 ANQKSKKVRSTIYLPDEFIRTLKGFKDGSPFEKQLHVRBANILMDDIGA----- 236
DB 923 -----LEYEKTIEDLKALQD-----ELKRNILM-DDISSYKQTTKL 961
QY 237 EEVTPWVRDEVIGLHLHYRAVHGLPTFPSSNPDY-----SLEHNL 277
DB 962 QEKIKWLER-----RSILIDELSEYSNQNPNLVOQKNELEERL 1004

```

Search completed: December 15, 2003, 15:14:17  
 Job time: 10.9023 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 15:11:40 ; Search time 36.1681 Seconds  
(without alignments)  
2233.384 Million cell updates/sec

Title: US-09-689-952-2

Perfect score: 1643  
Sequence: 1 MGCGGQIMQFKSIINTSD.....ERYKSLSTPYFLSGENFRNN 313

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

SPTREMBL\_23:\*  
1: sp archaea:\*  
2: sp bacteria:\*  
3: sp fungi:\*  
4: sp human:\*  
5: sp invertebrate:\*  
6: sp mammal:\*  
7: sp mhc:\*  
8: sp organelle:\*  
9: sp phage:\*  
10: sp plant:\*  
11: sp rodent:\*  
12: sp virus:\*  
13: sp vertebrate:\*  
14: sp unclassified:\*  
15: sp\_eukaryote:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1607	97.8	306	16	Q99TH8
2	1599	97.3	306	16	Q8NW67
3	1376	83.7	306	16	Q8CNV2
4	621.5	37.8	308	16	Q8CXB0
5	601.5	36.6	311	16	Q9K863
6	552.5	33.6	307	16	Q8Y6X1
7	547.5	33.3	307	16	Q92BP4
8	434.5	26.4	299	16	Q8CWM7
9	429.5	26.1	300	16	Q9A1D1
10	428.5	26.1	293	16	Q8P2J5
11	407.5	24.6	298	16	Q8CMP7
12	403.5	24.6	293	16	Q9CHM9
13	396	24.1	300	16	Q8B3T8
14	393	23.9	300	16	Q8DY72
15	392.5	23.9	298	16	Q97PC7
16	217	13.2	343	16	Q8EWK0

17	205.5	12.5	74	2	Q05653	005653 bacillus ce
18	173	10.5	209	2	Q48991	Q48991 mycoplasma
19	168.5	10.3	297	2	Q9R003	Q9R003 mycoplasma
20	159.5	9.7	259	9	Q9G022	Q9G022 bacteriophage
21	151.5	9.2	257	9	Q9B0F8	Q9B0F8 staphylococ
22	146	8.9	259	9	Q9MBR8	Q9MBR8 staphylococ
23	145	8.8	261	16	Q932A3	Q932A3 staphylococ
24	145	8.8	282	16	Q97HS5	Q97HS5 clostridium
25	145	8.8	316	16	Q9R014	Q9R014 ureaplasma
26	142.5	8.7	296	6	Q38098	Q38098 bacteriophage
27	138.5	8.4	235	16	Q67056	Q67056 aquifex ae
28	132	8.0	835	1	Q9V2X2	Q9V2X2 pyrococcus
29	131.5	8.0	329	16	Q97D92	Q97D92 clostridium
30	129	7.9	285	16	Q8E296	Q8E296 leptospira
31	128.5	7.8	739	16	Q51774	Q51774 borrelia bu
32	127	7.7	249	2	Q938D5	Q938D5 escherichia
33	127	7.7	587	17	Q962Y0	Q962Y0 sulfolobus
34	126.5	7.7	327	16	Q8R6U1	Q8R6U1 thermococcus
35	126	7.7	287	16	Q926A4	Q926A4 listeria in
36	121.5	7.4	396	5	Q62556	Q62556 manduca sex
37	120.5	7.3	316	2	P95446	P95446 pseudomonas
38	119	7.2	331	16	Q8XK72	Q8XK72 clostridium
39	118.5	7.2	446	2	Q8GQR4	Q8GQR4 bacillus we
40	118.5	7.2	488	16	Q8RDA3	Q8RDA3 thermococcus
41	117.5	7.2	267	16	Q9K732	Q9K732 bacillus ha
42	116	7.1	842	16	Q98Q49	Q98Q49 mycoplasma
43	114.5	7.0	584	17	Q96Y05	Q96Y05 sulfolobus
44	114	6.9	457	16	Q8XPG2	Q8XPG2 clostridium
45	113.5	6.9	195	17	Q8P7M8	Q8P7M8 methanobac

## ALIGNMENTS

### RESULT 1

Q99TH8 PRELIMINARY; PRT; 306 AA.  
ID Q99TH8  
AC Q99TH8  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Primosomal protein.  
GN DNAI OR SAVI684 OR SA1507.  
OS Staphylococcus aureus (Strain Mu50 / ATCC 700699), and  
OS Staphylococcus aureus (Strain N315).  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_Taxid=158878, 158879;  
[1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S. aureus (Strain Mu50), and S. aureus (Strain N315);  
RX MEDLINE=21311952; PubMed=11418146;  
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iino J.-O., Ito T.,  
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,  
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
RA Sekimizu K., Hirakawa H., Kohara S., Goto S., Yabuzaki J.,  
RA Kanihisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shida T.,  
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.,  
RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
RT aureus".  
RL Lancet 357:1225-1240(2001).  
DR EMBL; AP003134; BAB57846.1; -;  
DR EMBL; AP003134; BAB42774.1; -;  
KW Complete proteome.  
SQ SEQUENCE 306 AA; 35635 MW; 7F3440B89643505E CRC64;

Query Match 97.8%; Score 1607; DB 16; Length 306;  
Best Local Similarity 100.0%; Pred. No. 2.4e-115;  
Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 8 MGQFSIINTSDPERIKIKKEVINDDVQFLFAHRAELTNAMIDDLNVLQSYKQ 67  
DB 1 MGQFSIINTSDPERIKIKKEVINDDVQFLFAHRAELTNAMIDDLNVLQSYKQ 60

```

QY 68 QKHVDGHEKADCPNFKVGHVPELVYDNNRIKIRYLQCPCKIKYDEBERFEALITSHMOR 127
DB 61 QKHVDGHEKADCPNFKVGHVPELVYDNNRIKIRYLQCPCKIKYDEBERFEALITSHMOR 120
QY 128 DTLNAKLDIYMNHRDLVDVMAADDICTAITNGEQVGLYLPFGTGSFIIGAIANQ 187
DB 121 DTLNAKLDIYMNHRDLVDVMAADDICTAITNGEQVGLYLPFGTGSFIIGAIANQ 180
QY 188 LKSKKVRSTIIYLPFIRTLKGGFKDGSFEKKLHVRBANILMLDDIGAEVTPWVDEV 247
DB 181 LKSKKVRSTIIYLPFIRTLKGGFKDGSFEKKLHVRBANILMLDDIGAEVTPWVDEV 240
QY 248 IGPLHRYMHELPFPSSNFDYSELHHLAMTRDGEKTKAARIIRVYSLSTPYFLSG 307
DB 241 IGPLHRYMHELPFPSSNFDYSELHHLAMTRDGEKTKAARIIRVYSLSTPYFLSG 300
QY 308 ENFRNN 313
DB 301 ENFRNN 306

```

## RESULT 2

```

Q8NM67 ID Q8NM67 PRELIMINARY; PRT; 306 AA.
AC Q8NM67;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DE 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Primosomal protein.
GN DNAI OR MW1627.
OS Staphylococcus aureus (strain MW2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
RX NCBI_TaxID=19620;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwama N., Asano K., Naito T., Kuroda H., Chai L.,
RA Yamamoto K., Hitamatsu K.,
RT "Genome and virulence determinants of high virulence community-
RT acquired MRSA."
RL Lancet 359:1819-1827 (2002).
DR EMBL; AP004827; BAB95492.1; -.
KW Complete proteome.
SQ SEQUENCE 306 AA; 35626 MW; CQFCA3752D934B4 CRC64;

```

```

Query Match 97.3%; Score 1599; DB 16; Length 306;
Best Local Similarity 99.7%; Pred. No. 1e-114;
Matches 305; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 8 MKQPKSIINTSODEPKRIEIKKEVINDDPVKQFLNHRALTNAMIDEDLNTLOEYKQ 67
DB 1 MKQPKSIINTSODEPKRIEIKKEVINDDPVKQFLNHRALTNAMIDEDLNTLOEYKQ 60
QY 68 QKHVDGHEKADCPNFKVGHVPELVYDNNRIKIRYLQCPCKIKYDEBERFEALITSHMOR 127
DB 61 QKHVDGHEKADCPNFKVGHVPELVYDNNRIKIRYLQCPCKIKYDEBERFEALITSHMOR 120
QY 128 DTLNAKLDIYMNHRDLVDVMAADDICTAITNGEQVGLYLPFGTGSFIIGAIANQ 187
DB 121 DTLNAKLDIYMNHRDLVDVMAADDICTAITNGEQVGLYLPFGTGSFIIGAIANQ 180
QY 188 LKSKKVRSTIIYLPFIRTLKGGFKDGSFEKKLHVRBANILMLDDIGAEVTPWVDEV 247
DB 181 LKSKKVRSTIIYLPFIRTLKGGFKDGSFEKKLHVRBANILMLDDIGAEVTPWVDEV 240
QY 248 IGPLHRYMHELPFPSSNFDYSELHHLAMTRDGEKTKAARIIRVYSLSTPYFLSG 307
DB 241 IGPLHRYMHELPFPSSNFDYSELHHLAMTRDGEKTKAARIIRVYSLSTPYFLSG 300
QY 308 ENFRNN 313
DB 301 ENFRNN 306

```

```

DB 301 ENFRNN 306

```

## RESULT 3

```

Q8CNY2 ID Q8CNY2 PRELIMINARY; PRT; 306 AA.
AC Q8CNY2;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Primosomal protein.
GN SR1358.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
RX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
RA Chen Z., Wen Y.,
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB016748; AAC04957.1; -.
KW Complete proteome.
SQ SEQUENCE 306 AA; 35736 MW; EDD9F7B52BFC16 CRC64;

```

```

Query Match 83.7%; Score 1376; DB 16; Length 306;
Best Local Similarity 83.0%; Pred. No. 1.3e-97;
Matches 254; Conservative 33; Mismatches 19; Indels 0; Gaps 0;

```

```

QY 8 MKQPKSIINTSODEPKRIEIKKEVINDDPVKQFLNHRALTNAMIDEDLNTLOEYKQ 67
DB 1 MKQPKSIINTSODEPKRIEIKKEVINDDPVKQFLNHRALTNAMIDEDLNTLOEYKQ 60
QY 68 QKHVDGHEKADCPNFKVGHVPELVYDNNRIKIRYLQCPCKIKYDEBERFEALITSHMOR 127
DB 61 QKHVDGHEKADCPNFKVGHVPELVYDNNRIKIRYLQCPCKIKYDEBERFEALITSHMOR 120
QY 128 DTLNAKLDIYMNHRDLVDVMAADDICTAITNGEQVGLYLPFGTGSFIIGAIANQ 187
DB 121 DTLNAKLDIYMNHRDLVDVMAADDICTAITNGEQVGLYLPFGTGSFIIGAIANQ 180
QY 188 LKSKKVRSTIIYLPFIRTLKGGFKDGSFEKKLHVRBANILMLDDIGAEVTPWVDEV 247
DB 181 LKSKKVRSTIIYLPFIRTLKGGFKDGSFEKKLHVRBANILMLDDIGAEVTPWVDEV 240
QY 248 IGPLHRYMHELPFPSSNFDYSELHHLAMTRDGEKTKAARIIRVYSLSTPYFLSG 307
DB 241 IGPLHRYMHELPFPSSNFDYSELHHLAMTRDGEKTKAARIIRVYSLSTPYFLSG 300
QY 308 ENFRNN 313
DB 301 ENFRNN 306

```

## RESULT 4

```

Q8CXB0 ID Q8CXB0 PRELIMINARY; PRT; 308 AA.
AC Q8CXB0;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Primosomal protein (Helicase loader).
GN DNAI OR OB2156.
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacillales; Oceanobacillus.
RX NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HTE831 / DSM 14371 / JCM 11309;
RX MEDLINE=2220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.,
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme

```

RT environments.  
 RL Nucleic Acids Res. 30:3927-3935 (2002).  
 DR EMBL: AP004600; BAC14112.1; -  
 KW Helicase; Complete proteome.  
 SQ SEQUENCE 308 AA; 35729 MW; 8D4CF70E480A833 CRC64;

Query Match 37.8%; Score 621.5; DB 16; Length 308;  
 Best Local Similarity 41.4%; Pred. No. 9.5e-40;  
 Matches 127; Conservative 67; Mismatches 110; Indels 3; Gaps 3;

QY 7 IMQFSSINTSDPFERIKIKKIVINDPVKQFLAHRALTNAMIDEDLVLOEYKD 66  
 DB 4 IOSELKAMKKNKFFQTSYQNIKREKVIDPSIKHFL-ANPELISEIIDKHLITLYEYKS 62  
 QY 67 QOQHVDHK-FADCPNVKGVHPELVYDNNRIKIRYLCPCPKIKYDERFEALITSHM 125  
 DB 63 QSQKQDCCKSLGGQNNIOGSPVLEADNNDIRISTYKCHRLSEEQOQKLIQSLYM 122  
 QY 126 QBDTLNAKLDIYNNHRDLVMAADICTAITNGEYVGLYLPFGTGKSPILGATA 185  
 DB 123 PKHILQARISDVQDEH-RGNALGKVLDFLEASKEQLPKGLYIGSFVGKTYLGAIA 181  
 QY 186 NQLKSKKVRSTIYLPFIRTLKGFQSGFEKKLHVRANILMLDDIGAEVTPWVRD 245  
 DB 182 NELKLEYSISLYMPEFVEIKSSFQDSFNEKVDFFKADILMLDMGAEQMSAMFRD 241  
 QY 246 EVLGPILHYRMVHELPTFFSSNPDYSLEHHLMTREGEKTKAARITTEVKSISTPYL 305  
 DB 242 EVLASVLYQMMBGLPFTSNYDLDQLKQELSTTRNGEVQVAGRITIRIKOYTTDKL 301  
 QY 306 SCENFRN 312  
 DB 302 SGPNRRS 308

RESULT 5  
 QY 09K863 PRELIMINARY; PRT; 311 AA.  
 AC 09K863;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Primosome component (Helicase loader).  
 GN DNAI OR BH3144.  
 OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=86665;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RC MEDLINE=20512582; PubMed=11058132;  
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Maui N.,  
 RA Fujii F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
 RA Horikoshi K.;  
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
 RT halodurans and genomic sequence comparison with Bacillus subtilis."  
 RL Nucleic Acids Res. 28:4317-4331(2000).  
 DR EMBL: AP001517; BAB06863.1; -  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR SMART: SM00382; AAA; 1.  
 KM Helicase; Complete proteome.  
 SQ SEQUENCE 311 AA; 35822 MW; 6AFB3073431363BB CRC64;

Query Match 36.6%; Score 601.5; DB 16; Length 311;  
 Best Local Similarity 39.2%; Pred. No. 3.3e-38;  
 Matches 125; Conservative 69; Mismatches 104; Indels 21; Gaps 8;

QY 5 OSIMKQKSIINSDPFERIKIKKIVINDPVKQFLAHRALTNAMIDEDLVLOEYKD 64  
 DB 2 ESIQSSIKHMTG-NRSEKQPTOLKEAVFRSPHVLTEER-PELSPTTLEQGSKIYEX 59  
 QY 65 KDOQKHYDGH-KFADCPNVKGVHPELVYDNNRIKIRYLCPCPKIKYDERFEALIT 122

DB 60 QKEQSHC-AHCPGLQKCPNLMKGYPTLYVERDLSYSPCLKEEKEKKRSLIRS 118  
 QY 123 HHMQRTLVNAKLDI-----YNNHRDLVMAADICTAITNGEYVGLYLPFGT 175  
 DB 119 LVPKELTEKFDVSEPERGSTASHR-ALEPLASKP-----GSDGKGLYLPFGV 170  
 QY 176 GKSPILGATAINQSKSKVSTIYLPFIRTLKGFQSGFEKKLHVRANILMLDDIG 235  
 DB 171 GKTFLMGALINELKMDGIDSTIYVDFPRLKQSIGDGFQOKLPVKNQAVLIPDID 230  
 QY 236 ABEFTVWVDEVIQPLHYRMVHELPTFFSSNPDYSLEHHLMT-RDGEKTKAARIT 294  
 DB 231 AETMTSVWVDVGVILQYRIMEKLPFTLSNYDDELBEHLAYNDSGTLLKAKVME 290  
 QY 295 RVKSLTPYFLSGENFRN 313  
 DB 291 RINHYTVSWVQGNRRH 309

RESULT 6  
 ID 08Y6X1 PRELIMINARY; PRT; 307 AA.  
 AC 08Y6X1;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Primosome component (Helicase loader) Dnai.  
 GN DNAI OR LMO1560.  
 OS Listeria monocytogenes.  
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
 OX NCBI\_TaxID=1639;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=BGD-e / Serovar 1/2a;  
 RC MEDLINE=21537279; PubMed=11679669;  
 RA Glasier P., Frangoul L., Buchrieser C., Rusniok C., Amend A.,  
 RA Baquero F., Berche P., Bioecker H., Brandt P., Chakraborty T.,  
 RA Charbit A., Chetaniani F., Couve E., de Daruvar A., Deloux P.,  
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
 RA Entian K.-D., Fahn H., Garcia-del Portillo F., Garrido P.,  
 RA Gautier L., Gobeil W., Gomez-Lopez N., Hain T., Haut J., Jackson D.,  
 RA Jones L.-M., Kaerest U., Kreft J., Kuhn M., Kunst F., Kuyuphat G.,  
 RA Meduno E., Maitournan A., Mata Vicente J., Ng E., Nedjari H.,  
 RA Nordstok G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
 RA Rammel B., Rose M., Schlueter T., Simoes N., Tietze A.,  
 RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;  
 RT "Comparative genomics of Listeria species."  
 RL Science 294:849-852(2001).  
 DR EMBL: AL591979; CAC99638.1; -  
 DR Listliet; LMO01560; -  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR SMART: SM00382; AAA; 1.  
 KM Helicase; Complete proteome.  
 SQ SEQUENCE 307 AA; 35970 MW; CF2FD9350CEB0F48 CRC64;

Query Match 33.6%; Score 552.5; DB 16; Length 307;  
 Best Local Similarity 37.8%; Pred. No. 1.9e-34;  
 Matches 113; Conservative 69; Mismatches 104; Indels 13; Gaps 4;

QY 19 ODFEKRIEKIKKIVINDPVKQFLAHRALTNAMIDEDLVLOEYKDQKHYDGHFAD 78  
 DB 15 RDPKRYQGLKQGVHAYQPIQDPFKHKEVTEQLVQNMNSNYEWTQHKKTBE--- 71  
 QY 79 CPNFVKGVHPELVYDNNRIKIRYLCPCPKIKYDERFEALITSHMQRTLVNAKLDIY 138  
 DB 72 -ETLMGVAPKVLVINEPFTVYYPKEKIEBKRAVARRRISLTPMPQVADANADFY 130  
 QY 139 MNNHRDLVMAADICTAITN-----GROVGLYLPFGTGKSPILGATAINQSKKY 193  
 DB 131 TDESR---QLALVEYQFLANNYPKSGERVKGLFTHSGIGKSYLGLANKEMLKGI 187  
 QY 194 RSTIYLPFIRTLKGFQSGFEKKLHVRANILMLDDIGAEVTPWVRDEVIQPLH 253

```

Db      188 STLVLVLPFRFMRREVKOSISDNTVGEKIQPAKETEVLMMDIGAESMTAWTDEVLGAILQ 247
Qy      254 YRMVHELPTFSSNPDSELEHILAMTRDG--EETKTAIRIERKSLSTPFLSGENR 311
      248 FRMQEELPTFSSNPDSELEHILAMTRDG--EETKTAIRIERKSLSTPFLSGENR 306

RESULT 7
Q92BF4 PRELIMINARY; PRT; 307 AA.
AC Q92BF4;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Primosome component (Helicase loader) DnaI.
GN DnaI OR LIN1595.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / Serovar 6a;
RX PubMed=11679669;
RA Glaeser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
BA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Cherouani F., Couve E., de Daruvar A., Deboux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Ertian K.-D., Fehli H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Haut J., Jackson D.,
RA Jones L.-M., Kaerst U., Krefit J., Kuhn M., Kunst F., Kutapkat G.,
RA Madueno E., Maitouram A., Mata Vicente J., Ng E., Nedjari H.,
RA Nodjlek G., Novella S., de Pablos B., Perez-Diaz J.-C., Pucell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Coesart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL596169; CAC96826.1; -.
DR ListList; LIN01595; -.
DR InterPro; IPR003593; AAA_ATPase.
DR SMART; SM00382; AAA_1.
KW Helicase; Complete proteome.
SQ SEQUENCE 307 AA; 3598 MW; A651763BA8F123F CRC64;

Query Match 33.3%; Score 547.5; DB 16; Length 307;
Best Local Similarity 37.8%; Pred. No. 4.5e-34;
Matches 112; Conservative 68; Mismatches 109; Indels 7; Gaps 3;

Qy      19 QDFEKRIEIKGEYINDPVKQFLFAHRAELTNAMIDEDLVLOEYKDOQKRYDGHKPAD 78
      15 RDFEKRYQGLKQVLAHYQPIQDFKENKEDITEQLINONLSNIYEFMTQHKKFTGE--- 71
Db      79 CPNVKGVHVELYVDNNRIKIRYIQCPCKIKYDERFEALITSHNQRTIAKLDIY 138
      72 -ETLMPGVAPLVLNGEFITVYTPYRKIEEDRRVVERIRIRSLYMPKQVDMANLDFY 130
Qy      139 MNHEDRLDVMAADDICTAI--TNGEYVKGILYGPFGTGSFLIGAIANOLSKSVYST 196
      131 TDBESRKALVAEYQFLANNTPQNGEYKSLFTIGSGTGKSYILGALAKELAKIGISTT 190
Qy      197 IYVPEIRTLKGGFKDGSFEKKLHRYREANILMLDDIGAEVTPWVRDEVYGLLHYRM 256
      191 LVYLPFRMRREVKQISDNTVGEKIQPAKETEVLMMDIGAESMTAWTDEVLGAILQFRM 250
Qy      257 VHELPFTFSSNPDSELEHILAMTRDG--EETKTAIRIERKSLSTPFLSGENR 311
      251 QEBELPTFSSNPDSELEHILAMTRDG--EETKTAIRIERKSLSTPFLSGENR 306

RESULT 8
Q8CWM7 PRELIMINARY; PRT; 299 AA.
AC Q8CWM7;
DT 01-MAR-2003 (TReMBLrel. 23, Created)

```

```

DT      01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT      01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Putative DNA replication protein, primosome component (Helicase
DE loader).
GN DnaI OR SMU.1921.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UA159 / ATCC 700610 / Serotype C;
RX MEDLINE=2295063; PubMed=12397186;
RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
DR EMBL; AE015016; AAN59532.1; -.
KW Helicase; Complete proteome.
SQ SEQUENCE 299 AA; 34410 MW; 98CAAB56F5524BD1 CRC64;

Query Match 26.4%; Score 434.5; DB 16; Length 299;
Best Local Similarity 32.0%; Pred. No. 2e-25;
Matches 99; Conservative 68; Mismatches 127; Indels 15; Gaps 5;

Qy      4 GGSIMQFQSIINTSDFEKRIKIKYEVINDPVKQFLFAHRAELTNAMIDEDLVLOE 63
      5 GGTIAKDKNKHXYT-----EKMVALISDDEINFLAAH--HLSKQIKSLRFNQ 54
Db      64 YKQOQKHVDHKKPADCPNFKVGHVPELYVDNNRIKIRYIQCPCKIKYDERFEALITSH 123
      55 YRLERRFRFNH--DQATYIAKQPLVLMWEGADVAYKETKELIAKKTQALSDRINV 111
Qy      124 HMQRDTLNAKLDIYNHEDRLDVMAADDICTAITNGEYVKGILYGPFGTGSFLGA 163
      112 SLPRSYKNISFDINDVRLDVFVRVADVFVQYENPEQ--KGLYLYGDMGIGKSYLMAA 170
Db      184 IANQLSKK-VRSITTYLPFRTTLKGGFKDGSFEKKLHRYREANILMLDDIGAEVTPW 242
      171 MAHELSEORGAATTLHPFSFTIDVKNAINTVGKKEIDAVKADILLIIDDIGAEOSTSW 230
Qy      243 VREBVGPLHYVMVELPFFSSNPDSELEHILAMTRDGEEKTAIRIERKSLSTP 302
      231 IRBEVAVLIQYMLBELPFTFSNYSFDLEKALNIKSGSDTQAKVMEIRIYLAKE 290
Db      303 YFLSGENR 311
      291 IHLEGENR 299

RESULT 9
Q9A1D1 PRELIMINARY; PRT; 300 AA.
AC Q9A1D1;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Putative primosome component (Helicase loader).
GN DnaI OR SPY0340 OR SPYM3.0248.
OS Streptococcus pyogenes, and
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314, 198466;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.pyogenes; STRAIN=SF370 / ATCC 700294 / Serotype M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic G., Lyon K.,
RA Primeaux C., Szate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,

```

RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;  
 RT "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*,"  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES= *S. pyogenes*; STRAIN=MGAS315 / Serotype M3;  
 RX MEDLINE=22133808; PubMed=12122206;  
 RA Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,  
 RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,  
 RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,  
 RA Schlievert P.M., Musser J.M.,  
 RT Genome sequence of a serotype M3 strain of group A *Streptococcus*:  
 RT phage-encoded toxins, the high-virulence phenotype, and clone  
 RT emergence.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).  
 DR EMBL: AEO06498; AAK33392.1; -  
 KW Hypothetical protein; Helicase; Complete proteome.  
 SQ SEQUENCE 300 AA; 34146 MW; 489827A789338C5A CRC64;  
 Query Match 26.1%; Score 429.5; DB 16; Length 300;  
 Best Local Similarity 31.2%; Pred. No. 4.9e-25;  
 Matches 94; Conservative 74; Mismatches 126; Indels 7; Gaps 4;  
 QY 12 KSIIINSDPEKRIEIKKEVINDPVKQFLHRAELTNMIDEDLVLOEKYDOCKHY 71  
 DB 6 ETMAKLGQNTNRVSDQIOTIADPEVASISQH--HLSQEQINLSKNOFLVEROKY 63  
 QY 72 DGHKPADCPNFVGVHVELVYNNRIKIRYLOCPCKIKYDEERFEALITSHMQDPTLN 131  
 DB 64 ---QLNDPSYIAGKYOPILAMNEGADVSYLFEKELVEAKQAISRIQLVELPKSYRH 120  
 QY 133 AKLKDVIAMHRDLVDAMADDICTAITNGEYKGLYLPFGTSGFLIGALANOLSKR 191  
 DB 121 IHLSDIDVNNASWEAPSAILDPEVOYPSAQ--KGILYLDGMDIGKSYLLAAVAHLSK 179  
 QY 192 K-VRSTIIYLPFRITLKGFGKDSFEKRLHRYREANIIMLDIGABEVTPTVRDEVIGP 250  
 DB 180 KGVSTLLHFPSPFAIDVKNKASVKEBIDAKNPVILDDIGABQATSWRDEVLYQ 239  
 QY 251 LHYRMVHELPTFPSSNPDVSELEHHLAMTRDSEKTKAARIERYKSLTPFSLGENP 310  
 DB 240 ILQYRMLEELPTFPSTNSYSPADLERKWKATIKGSDETWQAKRVMERVYLARFHLGANKR 239  
 QY 311 R 311  
 DB 300 R 300  
 RESULT 10  
 O8P235 PRELIMINARY; PRT; 293 AA.  
 AC Q8P235;  
 DT 01-OCT-2002 (TREMBLrel. 22, Created)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
 DE Putative primosome component (helicase loader).  
 GN SPM18\_0332.  
 OS *Streptococcus pyogenes* (serotype M18).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 NCBI\_TaxId=186103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MGAS8232 / Serotype M18;  
 RX MEDLINE=21927593; PubMed=11917108;  
 RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,  
 RA Sylva G.L., Sturdevant D.E., Rickelers S.M., Porcella S.F.,  
 RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,  
 RA Kapur V., Daily J.A., Veasy L.G., Musser J.M.,  
 RT "Genome sequence and comparative microarray analysis of serotype M18  
 RT group A *Streptococcus* strains associated with acute rheumatic fever  
 RT outbreaks,"

RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).  
 DR EMBL: AEO09978; AAL97087.1; -  
 KW Helicase; Complete proteome; Hypothetical protein.  
 SQ SEQUENCE 293 AA; 33357 MW; 0E87AD94BDB599B6 CRC64;  
 Query Match 26.1%; Score 428.5; DB 16; Length 293;  
 Best Local Similarity 32.4%; Pred. No. 5.7e-25;  
 Matches 93; Conservative 70; Mismatches 117; Indels 7; Gaps 4;  
 QY 26 EKKKEVINDPVKQFLHRAELTNMIDEDLVLOEKYDOCKHYDGHKPADCPNFVG 85  
 DB 13 DQIOTIADPEVASISQH--HLSQEQINLSKNOFLVEROKY---QLNDPSYIAG 67  
 QY 86 HYPELVYNNRIKIRYLOCPCKIKYDEERFEALITSHMQDPTLNAKLKDVIAMHRDL 145  
 DB 68 YQIPLAMNEGADVSYLFEKELVEAKQAISRIQLVELPKSYRHILSDIDVNNASVKR 127  
 QY 146 DVAMADDICTAITNGEYKGLYLPFGTSGFLIGALANOLSKK-VRSTIIYLPFR 204  
 DB 128 EAPSAILDPEVOYPSAQ--KGILYLDGMDIGKSYLLAAVAHLSKGVSTLLHFPFA 186  
 QY 205 RTLKGFGKDSFEKRLHRYREANIIMLDIGABEVTPTVRDEVIGPLHYRMVHELPTFP 264  
 DB 187 IDVKNAISNGSVKEBIDAKNPVILDDIGABQATSWRDEVLYQLQYRMLEELPTFP 246  
 QY 265 SGNPDVSELEHHLAMTRDSEKTKAARIERYKSLTPFSLGENP 311  
 DB 247 TSNYSFADLERKWKATIKGSDETWQAKRVMERVYLARFHLGANKR 293  
 RESULT 11  
 O8CMP7 PRELIMINARY; PRT; 298 AA.  
 AC O8CMP7;  
 DT 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Primosome component (Helicase loader).  
 GN DNAT OR SPM155.  
 OS *Streptococcus pneumoniae* (strain ATCC BAA-295 / R6).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 NCBI\_TaxId=171101;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21429245; PubMed=11544234;  
 RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgelt S.,  
 RA DeHoff B.S., Batrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,  
 RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,  
 RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsumura P.,  
 RA McArthur S.M., McHenry M., McLeaster K., Mundy C.W., Nicase T.I.,  
 RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rokey P.,  
 RA Sun P.-M., Winkler M.B., Yang Y., Young-Bellido M., Zhao G.,  
 RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostock P.R. Jr., Skatrud P.L.,  
 RA Glass J.I.,  
 RT "Genome of the bacterium *Streptococcus pneumoniae* strain R6,"  
 RL J. Bacteriol. 183:5709-5717(2001).  
 DR EMBL: AEO08523; AAL00359.1; -  
 KW Helicase; Complete proteome.  
 SQ SEQUENCE 298 AA; 34574 MW; 3EA8DBD1DED8F22 CRC64;  
 Query Match 24.8%; Score 407.5; DB 16; Length 298;  
 Best Local Similarity 32.0%; Pred. No. 2.4e-23;  
 Matches 94; Conservative 64; Mismatches 121; Indels 15; Gaps 7;  
 QY 21 FKKRIEIKKEVINDPVKQFLHRAELTNMIDEDLVLOEKY-KDOCKHYDGHKPADCP 79  
 DB 17 YQDLYQKIMK---DPDVAATVQ--QESLNGDELRKSRISKFMQYITERKFKRG----DT 66  
 QY 80 PNFVGVHVELVYNNRIKIRYLOCPCKIKYDEERFEALITSHMQDPTLNAKLKDVIY 139  
 DB 67 DYIAGKYRILWMHNGADVSYETPELIIAKEKAIIKRLMLINPSSLKVVSPADYVR 126

QY 140 NHRRLDVAAMAADICTAITNGEQVGLYLYPGFTGKSPFLGAIANQLSKK-VRSSTII 198  
 DB 127 DDVGRLLVLMKMIFFVNDYPR--NLKGLYLGDFGVGSFVVALADLSKRVSSITLL 184  
 QY 199 YLPEFITLKGFGDSEFEKKLHVRREANIMLDDIGAEVTPVWDEVIGPLHYRMVH 258  
 DB 185 HYPFVIVDVKNALSDGVNKTLYVDIKSEVLLIDIGAEOSTVWVRDEIIQVILIQYMOE 244  
 QY 259 ELPPFSSNPDSYSELHHLAMTRDGEKKT--KAARIIKRVSLSTPYLSENGFR 311  
 DB 245 NLPTFTSNFNEFDELRHFAKVKFGNDETWEARRVMERIRYLAETRLBEGVNR 298

## RESULT 12

Q9CHN9 PRELIMINARY; PRT; 293 AA.  
 AC Q9CHN9  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Primosomal protein DnaI.  
 GN DnaI OR L10752.  
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.  
 NC NCB1\_TaxID=1360;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IL1403;  
 RX MEDLINE=2125186; PubMed=11337471;  
 RA Bolotin A., Winkler P., Manger S., Jallion O., Malarne K.,  
 RA Weisenbach J., Ehrlich S.D., Sorokin A.;  
 RT "The complete genome sequence of the lactic acid bacterium Lactococcus  
 RT lactis ssp. lactis IL1403.";  
 RL Genome Res. 11:731-753(2001).  
 DR EMBL; AB006308; AKC04850.1; -  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR SMART; SM00382; AAA; 1.  
 KM Complete proteome.  
 SQ SEQUENCE 293 AA; 33904 MW; D75GCS30A69445B3 CRC64;

Query Match 24.6%; Score 403.5; DB 16; Length 293;  
 Best Local Similarity 32.3%; Pred. No. 4,7e-23;

Matches 101; Conservative 61; Mismatches 124; Indels 27; Gaps 8;

QY 8 MKQPKSIINTSODEKRIEIKKEVINDPVKQFLHRAELTNAMIDEDLVNLQEV-KD 66  
 DB 1 MESIGDLGKRQDIRRENFELVAEVLKNADVAQFIAGH--OMTSDEIORSYSKFEYVRE 58  
 QY 67 QOKHYDGHKFPADCPNFVKGHPVPELYVDNRIRIKIRYLOCPCKIKYDEERFAELITSH-- 124  
 DB 59 HEKPEKGEKKA-----ADGYPEVLIMHGVADVS--QTTELAAQQAQNLLRPNIT 110  
 QY 125 -MQRDTLNAKLKDIYNNHRDLVDVAMAADICTAITNGEQVGLYLYPGFTGKSPFLGA 183  
 DB 111 GLPBDLQVTLADIALDVORIKPYQALVDF--ITNPPKKGIYLYGDFGVGSFYLAA 167  
 QY 184 IANQLSKKRVSTIYLPFIRTLKGFKDGSEFEK--LHVRREANIMLDDIGAEVTP 240  
 DB 168 MAEELAKKGISTILLHPTFI-----SLDPEFNARVWNEIKASVQLVLDIGAEQNN 220  
 QY 241 PVWDEVIGPLHYRMVHELPTFPSSNPDYSELEHHLAMTRDGEKKTAAARIIKRVSL 300  
 DB 221 AMWRDSTLYQVILQHRMOENLPTFTSTSLRMBELLEGHLAEFKRADETMPARVWERVYILA 280  
 QY 301 TPFYLSGENFRNN 313  
 DB 281 EEMRLBEGTNRHD 293

## RESULT 13

Q8E3T8 PRELIMINARY; PRT; 300 AA.  
 ID Q8E3T8  
 AC Q8E3T8;

DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 GN DnaI OR GBS1668.  
 OS Streptococcus agalactiae (serotype III).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 NC NCB1\_TaxID=216495;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NEM316 / Serotype III;  
 RX MEDLINE=22242508; PubMed=12354221;  
 RA Glaser P., Rusniok C., Buchrieser C., Chevallier F., Frangeul L.,  
 RA Meadek T., Zouine M., Couve E., Laliou L., Poyart C., Trieru-Cuot P.,  
 RA Kunst F.;  
 RT "Genome sequence of Streptococcus agalactiae, a pathogen causing  
 RT invasive neonatal disease.";  
 RL Mol. Microbiol. 45:1499-1513(2002).  
 DR EMBL; AL766852; CAD47327.1; -  
 DR Segalier; gbs1668; -  
 KM Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 300 AA; 34679 MW; 9C0B4898661661B9 CRC64;

Query Match 24.1%; Score 396; DB 16; Length 300;  
 Best Local Similarity 28.5%; Pred. No. 1,8e-22;

Matches 88; Conservative 61; Mismatches 126; Indels 14; Gaps 5;

QY 4 GQSIMQFISINTSODEKRIEIKKEVINDPVKQFLHRAELTNAMIDEDLVNLQEV 63  
 DB 5 GQALENGQVRPRNTN-----DELQIMLSDQVAQVLEFKTH--QLSQREINISMSEKPNQ 55  
 QY 64 YKQOQHGHGKFPADCPNFVKGHPVPELYVDNRIRIKIRYLOCPCKIKYDEERFAELITSH 123  
 DB 56 FLIERKF--KKKQSYIAKGTEPLVNNEGVADVSYTELLEIAQKKQALSDRINLV 112  
 QY 124 HMQRDTLNAKLKDIYNNHRDLVDVAMAADICTAITNGEQVGLYLYPGFTGKSPFLGA 183  
 DB 113 NLPKSRNIRMTDFDINNBSRMKAMGQLDPFETYSVNH--KGLYLYGDMGVGKSYLMA 171  
 QY 184 IANQLSKK-VNSTIYLPFIRTLKGFKDGSEFEKKLHVRREANIMLDDIGAEVTPW 242  
 DB 172 MAEELSERKGVSTTLHFPSPFADVNAISSGVKDEIVAVSVPLIILDDIGAEQATSW 231  
 QY 243 VREDEVIGPLHYRMVHELPTFPSSNPDYSELEHHLAMTRDGEKKTAAARIIKRVSLSP 302  
 DB 232 VRDEIIQVILQHRMSELPTFTSTSLRMBELLEGHLAEFKRADETMPARVWERVYLAIE 291  
 QY 303 YPLSGENFR 311  
 DB 292 FHLBEGPNR 300

## RESULT 14

Q8DY72 PRELIMINARY; PRT; 300 AA.  
 ID Q8DY72  
 AC Q8DY72;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Primosomal protein DnaI.  
 GN DnaI OR SAG1621.  
 OS Streptococcus agalactiae (serotype V).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 NC NCB1\_TaxID=216466;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=2603 V/R / Serotype V;  
 RX MEDLINE=22222368; PubMed=12200547;  
 RA Tetelsh H., Masignani V., Cieslewicz M.J., Eisen J.A., Peterson S.,  
 RA Wessels M.R., Paulsen I.T., Nelson K.E., Margalit I., Read T.D.,  
 RA Madoff L.C., Wolf A.M., Beaman M.J., Brinkac L.M., Daugherty S.C.,

RA DeBoy R.T., Durkin A.S., Kojonay J.F., Madupu R., Lewis M.R.,  
 RA Radune D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S.,  
 RA Cary H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,  
 RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vogni F., Malone D.,  
 RA Rimando D., Rappunli R., Telford J.V., Kasper D.L., Grandi G.,  
 RA Fraser C.M.;  
 RT "Complete genome sequence and comparative genomic analysis of an  
 RT emerging human pathogen, serotype V *Streptococcus agalactiae*,"  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).  
 DR EMBL: AB014265; AAN00485.1; -.  
 DR TIGR: SAG1621; -.  
 KW Complete proteome.  
 SQ SEQUENCE 300 AA; 34663 MW; 9BED36F8B363B9 CRC64;  
 Query Match 23.9%; Score 393; DB 16; Length 300;  
 Best Local Similarity 28.5%; Pred. No. 3.1e-22;  
 Matches 88; Conservative 80; Mismatches 127; Indels 14; Gaps 5;  
 QY 4 GOSIMQFQSIINTSODFEKRIKKEKVEINDPVKQFLHRAELTNAMIDDLNVLOE 63  
 DB 5 GQALEHQGVPRRTN-----DELQIMILADQVAFETTH--QLSQREINISMKEKQ 55  
 QY 64 YKDQKHVDGKFKADCPNFKGVHVELYDNNRIKIRYLOCPKIKYDEERFEAEILTSH 123  
 DB 56 FLIERQKF--KKKDSQYIAKGYEPILVMNMGVADVSYLETRILBAQKKQAISSDRINLV 112  
 QY 124 HMQRDITLNAKLDKIYNNHRRLVDVMAADICTAINTGBOVKGLYIGPFGTKSFILGA 183  
 DB 113 NPKSYRNIRMTDFDINNBSRMKMSQLDFVETTPSYNH-KELYLYGDMGVKSYLMAA 171  
 QY 184 IANQLSKK-VRSITLYLPEFIRTLKGGFKDGSFEKKLHVRBANIMLDDIGAEVTPW 242  
 DB 172 MARELSERKGVSTTLHFPESFALDVKAISSTGYKDEIDAVKSPILIIDDIGAEQATSW 231  
 QY 243 VRDEVIGPLHYMVELPTFPSSNPDYSLEHHLAMTRDGEKTKAARILIERVKSLSLP 302  
 DB 232 VRDEILQVILQHRMLELPTFPSTNSYSPNDLERKMANIKGSDTQAKRVMERRYLAIE 291  
 QY 303 YPLSGENR 311  
 DB 292 FHLEGNRR 300  
 RESULT 15  
 Q97PC7 PRELIMINARY; PRT; 298 AA.  
 ID Q97PC7; AC Q97PC7; DT 01-OCT-2001 (T-EMBLrel. 18, Created)  
 DT 01-OCT-2001 (T-EMBLrel. 18, Last sequence update)  
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)  
 DE Primosomal protein DnaI.  
 GN SP1711.  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1313;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TIGR4;  
 RX MEDLINE=21357209; PubMed=11463916;  
 RA Tettelin H., Nelson K.B., Paulsen I.T., Bisen J.A., Read T.D.,  
 RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,  
 RA Durkin A.S., Gwinn M., Kojonay J.F., Nelson W.C., Peterson J.D.,  
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,  
 RA Holtzapple E., Khouri H., Wolf A.M., Uterback T.R., Hansen C.L.,  
 RA McDonald L.A., Feldblyum T.V., Angluoli S., Dickinson T., Hickey E.K.,  
 RA Holt I.B., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
 RT "Complete genome sequence of a virulent isolate of *Streptococcus*  
 RT pneumoniae,"  
 RL Science 293:498-506(2001).  
 DR EMBL: AB007464; AAK75789.1; -.  
 DR TIGR: SP1711; -.

KW Complete proteome.  
 SQ SEQUENCE 298 AA; 34259 MW; CAE5ECAFA9C30059 CRC64;  
 Query Match 23.9%; Score 392.5; DB 16; Length 298;  
 Best Local Similarity 32.3%; Pred. No. 3.3e-22;  
 Matches 96; Conservative 64; Mismatches 116; Indels 21; Gaps 8;  
 QY 21 FEKRIEIKKEVINDPVKQFLHRAELTNAMIDDLNVLOE-KDQKHVDGKFKADC 79  
 DB 17 YQDLVQKIMK---DDVNAFTQ--QESLTPKEINRSIKFQYTERKFLRG----DT 66  
 QY 80 PNFVKGVHVELYDNNRIKIRYLOCPKIKYDEERFEAEILTSHMQRDITLNAKLDIYM 139  
 DB 67 DYIAGYKPIVMNMGVADVSYLETRILBAKKEAIIKRLILNIPASLKKASLAQVLD 126  
 QY 140 NHRDRLDV--AMADDICTAINTGBOVKGLYIGPFGTKSFILGAINQLSKK-VRS 195  
 DB 127 DDLGRLPVPEKLLAFVEQYPAIR-----KGLYLYGDFGVGKSPMAALAHDLSEKRGVSS 181  
 QY 196 TIIYLPEFIRTLKGGFKDGSFEKKLHVRBANIMLDDIGAEVTPWVRDEVIGPLHYR 255  
 DB 182 TILHTPSFVIDVKMAISGNAVKTLYDEIKLSVLLIIDDIGAEOSTVWVRDEILQVILQYR 241  
 QY 256 MYHELPTFPSSNPDYSLEHHLAMTRDGEKTKAARILIERVKSLSLPYPLSGENR 311  
 DB 242 MQENLPTFPSTNPFNDELEKHFPAKVHGNDETWEARVWERIRYLAETRLLEGVNR 298  
 Search completed: December 15, 2003, 15:15:56  
 Job time : 39.1651 secs

**THIS PAGE BLANK (USPTO)**